



11-02-05

1121

PTO/SB/21 (09-04)

Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

		Application Number	10/721,922-Conf. #5830
		Filing Date	November 24, 2003
		First Named Inventor	Markus POMPEJUS
		Art Unit	1635
		Examiner Name	Not Yet Assigned
Total Number of Pages in This Submission		Attorney Docket Number	BGI-132CPCN

ENCLOSURES (Check all that apply)

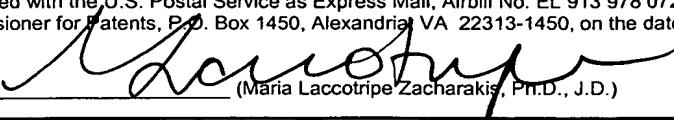
<input type="checkbox"/> Fee Transmittal Form	<input type="checkbox"/> Drawing(s)	<input type="checkbox"/> After Allowance Communication to TC
<input type="checkbox"/> Fee Attached	<input type="checkbox"/> Licensing-related Papers	<input type="checkbox"/> Appeal Communication to Board of Appeals and Interferences
<input type="checkbox"/> Amendment/Reply	<input type="checkbox"/> Petition	<input type="checkbox"/> Appeal Communication to TC (Appeal Notice, Brief, Reply Brief)
<input type="checkbox"/> After Final	<input type="checkbox"/> Petition to Convert to a Provisional Application	<input type="checkbox"/> Proprietary Information
<input type="checkbox"/> Affidavits/declaration(s)	<input type="checkbox"/> Power of Attorney, Revocation Change of Correspondence Address	<input type="checkbox"/> Status Letter
<input type="checkbox"/> Extension of Time Request	<input type="checkbox"/> Terminal Disclaimer	<input checked="" type="checkbox"/> Other Enclosure(s) (please identify below):
<input type="checkbox"/> Express Abandonment Request	<input type="checkbox"/> Request for Refund	Certificate of Express Mailing; Request for Corrected Patent Application Publication; copy of Patent Application Publication US 2005-0191732 A1 with changes marked; copy of Preliminary Amendment and Supplemental ADS filed with USPTO on March 24, 2004; Copy of Declaration, Petition and Power of Attorney; Return Postcard
<input type="checkbox"/> Information Disclosure Statement	<input type="checkbox"/> CD, Number of CD(s) _____	
<input type="checkbox"/> Certified Copy of Priority Document(s)	<input type="checkbox"/> Landscape Table on CD	
<input type="checkbox"/> Reply to Missing Parts/ Incomplete Application	Remarks	
<input type="checkbox"/> Reply to Missing Parts under 37 CFR 1.52 or 1.53		

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	LAHIVE & COCKFIELD, LLP		
Signature			
Printed name	Maria Laccotripe Zacharakis, Ph.D., J.D.		
Date	November 1, 2005	Reg. No.	56,266

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail, Airbill No. EL 913 978 072 US, in an envelope addressed to: MS PGPUB, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: November 1, 2005

Signature:  (Maria Laccotripe Zacharakis, Ph.D., J.D.)



Application No. (if known): 10/721,922

Attorney Docket No.: BGI-132CPCN

Certificate of Express Mailing Under 37 CFR 1.10

I hereby certify that this correspondence is being deposited with the United States Postal Service as Express Mail, Airbill No. EL 913 978 072 US in an envelope addressed to:

MS PGPUB
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

on November 1, 2005
Date


Signature

Maria Laccotripe Zacharakis, Ph.D., J.D.

Typed or printed name of person signing Certificate

56,266
Registration Number, if applicable

(617) 227-7400
Telephone Number

Note: Each paper must have its own certificate of mailing, or this certificate must identify each submitted paper.

This Certificate of Express Mailing (1 page)

Transmittal (1 page)

Request For Corrected Patent Application Publication Under 37 CFR 1.221(b) (2 pages)

Copy of Patent Application Publication US 2005/0191732 A1 with changes indicated at Page 1 (64 pages)

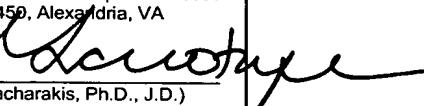
Copy of Preliminary Amendment and Supplemental Application Data Sheet filed with USPTO on March 24, 2004 (9 pages)

Copy of Declaration, Petition and Power of Attorney (7 pages)

Return Receipt Postcard

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail Airbill No. EL 913 978 072 US in an envelope addressed to: MS PGPub, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: November 1, 2005

Signature: 
(Maria Laccotripe Zacharakis, Ph.D., J.D.)

Docket No.: BGI-132CPCN
(PATENT)



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of: Markus Pompejus *et al.*

Application No.: 10/721,922

Confirmation No.: 5830

Filed: November 24, 2003

Art Unit: 1635

For: CORYNEBACTERIUM GLUTAMICUM GENES
ENCODING PROTEINS INVOLVED IN
HOMEOSTASIS AND ADAPTATION

Examiner: Not Yet Assigned

REQUEST FOR CORRECTED PATENT APPLICATION PUBLICATION UNDER 37 CFR § 1.221(b)

MS PGPUB
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Applicants hereby request that a corrected Patent Application Publication be issued in the above-identified patent application in accordance with 37 CFR § 1.221(b). The Patent Application Publication US 2005/0191732 A1 issued on September 1, 2005, a copy of which is attached hereto with changes marked on Page 1, has errors in the priority date of German Applications 19932914.1 and 19933006.9 located in the Foreign Application Priority Data Section and should be corrected as follows:

In the Foreign Application Priority Data Section, at Page 1

Jul. 9, 1999 (DE).....19932914.1 should be changed to: Jul. 14, 1999 (DE).....19932914.1; and

Jul. 9, 1999 (DE).....19933006.9 should be changed to: Jul. 14, 1999 (DE).....19933006.9

Applicants submit that the foregoing errors are material mistakes which are apparent from the Office records. This correction is supported by a copy of a Preliminary Amendment, together with a Supplemental Application Data Sheet, which were filed with the U.S. Patent and Trademark Office on March 24, 2004 in order to correct the errors in the priority dates of the aforementioned German Priority Applications for the above-identified patent application. In addition, Applicants submit a copy of the Declaration, Petition and Power of Attorney as submitted with the initial filing of this continuation application on November 24, 2003 which indicates the correct priority dates.

Applicants additionally request that all pertinent U.S. Patent and Trademark Office records relating to the subject application be changed to reflect this correction.

Applicants believe that no fee is due with this request. However, if a fee is due, please charge our Deposit Account No. 12-0080, under Order No. BGI-132CPCN from which the undersigned is authorized to draw.

Dated: November 1, 2005

Respectfully submitted,

By 
Maria Laccotripe Zacharakis, Ph.D., J.D.
Registration No.: 56,266
LAHIVE & COCKFIELD, LLP
28 State Street
Boston, Massachusetts 02109
(617) 227-7400
(617) 742-4214 (Fax)
Attorney/Agent For Applicant

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the application of: Markus Pompejus, et al.

Serial No.: 10/721922

Filed: November 24, 2003

For: "Corynebacterium Glutamicum Genes Encoding
Proteins Involved In Homeostasis And Adaptation"

Attorney Docket No.: BGI-132CPCN

Group Art Unit:

Examiner:



COPY

Mail Stop Non-Fee Amendment
Commissioner for Patents
Post Office Box 1450
Alexandria, VA 22313-1450

CERTIFICATION UNDER 37 CFR 1.10

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail Receipt No. EV 378 817 930 US in an envelope addressed to: Mail Stop Non-Fee Amendment, Commissioner for Patents, Post Office Box 1450, Alexandria, VA 22313-1450, on the date shown below.

March 24, 2004

Date of Signature and of Mail Deposit

By:

Lisa M. DiRocco,
Registration No. 51,619
Attorney for Applicants

PRELIMINARY AMENDMENT

Dear Sir:

Prior to examination of the above-identified application, please amend the application as follows:

Amendments to the Specification begin on page 3 of this paper

Remarks/Arguments begin on page 4 of this paper

Amendments to the Specification

Please replace the paragraph beginning at page 1, line 4, with the following replacement paragraph:

Related Applications

This application is a continuation application of U.S. Serial No. 09/603,124, filed on June 23, 2000, and This application claims priority to prior filed U.S. Provisional Patent Application Serial No. 60/141031, filed June 25, 1999, U.S. Provisional Patent Application Serial No. 60/143694, filed July 14, [[2000]] 1999, and U.S. Provisional Patent Application Serial No. 60/151778, filed August 31, 1999. This application also claims priority to German Application No. 19931418.7, filed July 8, 1999, German Application No. 19932124.8, filed July 9, 1999, German Application No. 19932126.4, filed July 9, 1999, German Application No. 19932127.2, filed July 9, 1999, German Application No. 19932133.7, filed July 9, 1999, German Application No. 19932207.4, filed July 9, 1999, German Application No. 19932208.2, filed July 9, 1999, German Application No. 19932225.2, filed July 9, 1999, German Application No. 19932229.5, filed July 9, 1999, German Application No. 19932914.1, filed July [[9]] 14, 1999, German Application No. 19933006.9, filed July [[9]] 14, 1999, German Application No. 19940765.7, filed August 27, 1999, German Application No. 19940768.1, filed August 27, 1999, German Application No. 19940831.9, filed August 27, 1999, German Application No. 19940832.7, filed August 27, 1999, German Application No. 19941385.1, filed August 31, 1999, German Application No. 19941396.7, filed August 31, 1999, and German Application No. 19942087.4, filed September 3, 1999. The entire contents of all of the aforementioned applications are hereby expressly incorporated herein by this reference.

Please replace the paragraph beginning at page 57, line 17, with the following replacement paragraph:

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see e.g., Fleischman, R.D. *et al.* (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., Science, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' (SEQ ID NO:479) or 5'-GTAAAACGACGCCAGT-3' (SEQ ID NO:480).

REMARKS

The specification has been amended to insert priority information in the first paragraph, and to correct typographical errors. The specification was also amended to insert sequence identifiers. No new matter has been added.

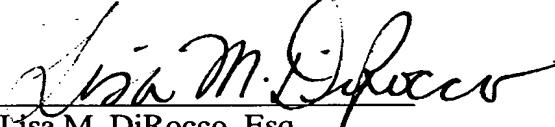
CONCLUSION

In view of the amendments and remarks set forth above, it is respectfully submitted that this application is in condition for allowance. If there are any remaining issues or the Examiner

believes that a telephone conversation with Applicants' Attorney would be helpful in expediting prosecution of this application, the Examiner is invited to call the undersigned at (617) 227-7400.

Respectfully submitted,

LAHIVE & COCKFIELD, LLP



Lisa M. DiRocco, Esq.
Reg. No. 51,619
Attorney for Applicants

28 State Street
Boston, MA 02109
Tel: (617) 227-7400

Dated: March 24, 2004



COPY

Supplemental Application Data Sheet

Application Information

Application number:: 10/721922
Filing Date:: 11/24/03
Application Type:: Regular
Subject Matter:: Utility
Suggested Group Art Unit:: N/A
CD-ROM or CD-R?:: None
Sequence submission?:: None
Computer Readable Form (CRF)?:: No
Title:: CORYNEBACTERIUM GLUTAMICUM
GENES ENCODING PROTEINS
INVOLVED IN HOMEOSTASIS AND
ADAPTATION
Attorney Docket Number:: BGI-132CPCN
Request for Early Publication?:: No
Request for Non-Publication?:: No
Small Entity?:: No
Petition included?:: No
Secrecy Order in Parent Appl.?:: No

Applicant Information

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Markus
Family Name:: POMPEJUS
City of Residence:: Waldsee
Country of Residence:: Germany
Street of mailing address:: Wenjenstrasse 21
City of mailing address:: Freinsheim

Country of mailing address:: Germany

Postal or Zip Code of mailing address:: 67251

Applicant Authority Type:: Inventor

Primary Citizenship Country:: Germany

Status:: Full Capacity

Given Name:: Burkhard

Family Name:: KRÖGER

City of Residence:: Limburgerhof

Country of Residence:: Germany

Street of mailing address:: Im Waldhof 1

City of mailing address:: Limburgerhof

Country of mailing address:: Germany

Postal or Zip Code of mailing address:: 67117

Applicant Authority Type:: Inventor

Primary Citizenship Country:: Germany

Status:: Full Capacity

Given Name:: Hartwig

Family Name:: SCHRÖDER

City of Residence:: Nussloch

Country of Residence:: Germany

Street of mailing address:: Goethestr. 5

City of mailing address:: Nussloch

Country of mailing address:: Germany

Postal or Zip Code of mailing address:: 69226

Applicant Authority Type:: Inventor

Primary Citizenship Country:: Germany

Status:: Full Capacity

Given Name:: Oskar

Family Name:: ZELDER

City of Residence:: Speyer
Country of Residence:: Germany
Street of mailing address:: Franz-Stutzel 8
City of mailing address:: Speyer
Country of mailing address:: Germany
Postal or Zip Code of mailing address:: 67346

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Gregor
Family Name:: HABERHAUER
City of Residence:: Limburgerhof
Country of Residence:: Germanv
Street of mailing address:: Moselstr. 42
City of mailing address:: Limburgerhof
Country of mailing address:: Germany
Postal or Zip Code of mailing address:: 67117

Correspondence Information

Correspondence Customer Number:: 00959

Representative Information

Representative Customer Number:: 00959

Domestic Priority Information

Application::	Continuity Type::	Parent Application::	Parent Filing Date::
This Application	Continuation of	09/603124	06/23/00
09/603124	An application claiming the benefit under 35 USC 119(e)	60/141031	06/25/99
09/603124	An application claiming the benefit under 35 USC 119(e)	60/143694	07/14/99
09/603124	An application claiming the benefit under 35 USC 119(e)	60/151778	08/31/99

Foreign Priority Information

Country::	Application number::	Filing Date::	Priority Claimed::
Germany	19931418.7	07/08/99	<u>Yes</u>
Germany	19932124.8	07/09/99	<u>Yes</u>
Germany	19932126.4	07/09/99	<u>Yes</u>
Germany	19932127.2	07/09/99	Yes
Germany	19932133.7	07/09/99	Yes
Germany	19932207.4	07/09/99	Yes
Germany	19932208.2	07/09/99	Yes
Germany	19932225.2	07/09/99	Yes
Germany	19932229.5	07/09/99	Yes
Germany	19932914.1	<u>07/14/99</u>	Yes
Germany	19933006.9	<u>07/14/99</u>	Yes
Germany	19940765.7	08/27/99	Yes
Germany	19940768.1	08/27/99	Yes
Germany	19940831.9	08/27/99	Yes
Germany	19940832.7	08/27/99	Yes
Germany	19941385.1	08/31/99	Yes
Germany	19941396.7	08/31/99	Yes
Germany	19942087.4	09/03/99	Yes

Assignee Information

Assignee name:: BASF Aktiengesellschaft
 Street of mailing address:: ZDZ/G
 City of mailing address:: Ludwigshafen
 Country of mailing address:: Germany
 Postal or Zip Code of mailing address:: D-67056



COPY

Customer Number: 000959

Attorney's
Docket
Number BGI-132CP

DECLARATION, PETITION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

**CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
INVOLVED IN HOMEOSTASIS AND ADAPTATION**

the specification of which:

_____ is attached hereto.

X was filed on June 23, 2000 as Application Serial No. 09/603,124
and was amended on _____
(if applicable)

I do not know and do not believe that the subject matter of this application was known or used by others in the United States or patented or described in a printed publication in any country before my invention thereof, or patented or described in a printed publication in any country or in public use or on sale in the United States more than one year prior to the date of this application, or first patented or caused to be patented or made the subject of an inventor's certificate by me or my legal representatives or assigns in a country foreign to the United States prior to the date of this application on an application filed more than twelve months (six months if this application is for a design) before the filing of this application; and I acknowledge my duty to disclose information of which I am aware which is material to the examination of this application, that no application for patent or inventor's certificate on the subject matter of this application has been filed by me or my representatives or assigns in any country foreign to the United States, except those identified below, and that I have reviewed and understand the contents of the specification, including the claims as amended by any amendment referred to herein.

I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

CLAIM OF BENEFIT OF EARLIER FOREIGN APPLICATION(S)

I hereby claim priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below, and have also identified below any foreign application(s) for patent or inventor's certificate filed by me on the same subject matter having a filing date before that of the application(s) from which priority is claimed.

Check one:

no such applications have been filed.
 such applications have been filed as follows

EARLIEST FOREIGN APPLICATION(S), IF ANY, FILED WITHIN 12 MONTHS
 (6 MONTHS FOR DESIGN) PRIOR TO THIS U.S. APPLICATION

Country	Application Number	Date of Filing (month,day,year)	Priority Claimed Under 35 USC 119
DE	19931418.7	07/08/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932124.8	07/09/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932126.4	07/09/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932127.2	07/09/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932133.7	07/09/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932207.4	07/09/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932208.2	07/09/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932225.2	07/09/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932229.5	07/09/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19932914.1	07/14/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19933006.9	07/14/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19940765.7	08/27/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19940768.1	08/27/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19940831.9	08/27/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19940832.7	08/27/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19941385.1	08/31/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19941396.7	08/31/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
DE	19942087.4	09/03/99	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

ALL FOREIGN APPLICATION(S), IF ANY FILED MORE THAN 12 MONTHS
(6 MONTHS FOR DESIGN) PRIOR TO THIS U.S. APPLICATION

CLAIM FOR BENEFIT OF U.S. PROVISIONAL APPLICATION(S)

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below.

60/141,031 June 25, 1999
(Application Serial No.) (Filing Date)

60/143,694 July 14, 1999
(Application Serial No.) (Filing Date)

60/151,778 August 31, 1999
(Application Serial No.) (Filing Date)

CLAIM FOR BENEFIT OF EARLIER U.S./PCT APPLICATION(S)

I hereby claim the benefit under Title 35, United States Code, §120 of any earlier United States application(s) or PCT international application(s) designating the United States listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the earlier application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which became available between the filing date(s) of the earlier application(s) and the national or PCT international filing date of this application. As to subject matter of this application which is common to my earlier application(s), if any, described below, I do not know and do not believe that the same was known or used by others in the United States or patented or described in a printed publication in any country before my invention thereof, or patented or described in a printed publication in any country or in public use or on sale in the United States more than one year prior to the date(s) of said earlier application(s), or first patented or caused to be patented or made the subject of an inventor's certificate by me or my legal representatives or assigns in a country foreign to the United States prior to the date(s) of said earlier application(s) on an application filed more than twelve months (six months if this application is for a design) before the filing of said earlier application(s); and I acknowledge that no application for patent or inventor's certificate on said subject matter has been filed by me or my representatives or assigns in any country foreign to the United States except those identified herein.

(Application Serial No.)	(Filing Date)	(Status) (patented,pending,aband.)
--------------------------	---------------	---------------------------------------

(Application Serial No.)	(Filing Date)	(Status) (patented,pending,aband.)
--------------------------	---------------	---------------------------------------

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorneys and/or agents to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

James E. Cockfield	Reg. No. 19,162	Jeanne M. DiGiorgio	Reg. No. 41,710
Thomas V. Smurzynski	Reg. No. 24,798	Megan E. Williams	Reg. No. 43,270
Ralph A. Loren	Reg. No. 29,325	Nicholas P. Triano III	Reg. No. 36,397
Giulio A. DeConti, Jr.	Reg. No. 31,503	Peter C. Lauro	Reg. No. 32,360
Ann Lampert Hammitt	Reg. No. 34,858	Timothy J. Douros	Reg. No. 41,716
Elizabeth A. Hanley	Reg. No. 33,505	DeAnn F. Smith	Reg. No. 36,683
Amy E. Mandragouras	Reg. No. 36,207	William D. DeVaul	Reg. No. 42,483
Anthony A. Laurentano	Reg. No. 38,220	David J. Rikkers	Reg. No. 43,882
Jane E. Remillard	Reg. No. 38,872	Chi Suk Kim	Reg. No. 42,728
Jeremiah Lynch	Reg. No. 17,425	Maria Laccotripe Zacharakis	Limited Recognition
Kevin J. Canning	Reg. No. 35,470	Debra J. Milasincic	Under 37 C.F.R. § 10.9(b)
David A. Lane, Jr.	Reg. No. 39,261	David R. Burns	Reg. No. P46,931
Catherine J. Kara	Reg. No. 41,106		Reg. No. P46,590

Send Correspondence to Giulio A. DeConti, Jr., Esq. at Customer Number: 000959 whose address is:

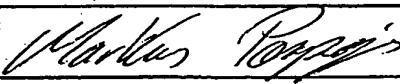
Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109

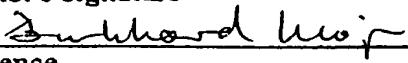
Direct Telephone Calls to: (name and telephone number)

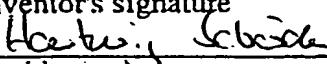
Giulio A. DeConti, Jr., Esq., (617) 227-7400

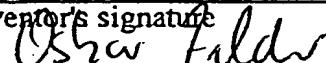
Wherefore I petition that letters patent be granted to me for the invention or discovery described and claimed in the attached specification and claims, and hereby subscribe my name to said specification and claims and to the foregoing declaration, power of attorney, and this petition.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor Markus Pompejus	Date
Inventor's signature 	July 06, 2000
Residence Wenjenstrasse 21, 67251 Freinsheim, Germany	
Citizenship Germany	
Post Office Address (if different)	

Full name of second inventor, if any Burkhard Kröger	
Inventor's signature 	Date July 06, 2000
Residence Im Waldhof 1, 67117 Limburgerhof, Germany	
Citizenship Germany	
Post Office Address (if different)	

Full name of third inventor, if any Hartwig Schröder	
Inventor's signature 	Date July 06, 2000
Residence Goethestr. 5, 69226 Nussloch, Germany	
Citizenship Germany	
Post Office Address (if different)	

Full name of fourth inventor, if any Oskar Zelder	
Inventor's signature 	Date July 06, 2000
Residence Rossmarktstr. 27, 67346 Speyer, Germany	
Citizenship Germany	
Post Office Address (if different)	

Full name of fifth inventor, if any Gregor Haberhauer	
Inventor's signature 	Date July 06, 2000
Residence Moselstr. 42, 67117 Limburgerhof, Germany	
Citizenship Germany	
Post Office Address (if different)	



US 20050191732A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2005/0191732 A1
(43) Pub. Date: Sep. 1, 2005

(54)	CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN HOMEOSTASIS AND ADAPTATION	Jul. 9, 1999 (DE).....	19932133.7
		Jul. 9, 1999 (DE).....	19932207.4
		Jul. 9, 1999 (DE).....	19932208.2
		Jul. 9, 1999 (DE).....	19932225.2
(75)	Inventors: Markus Pompejus, Waldsee (DE); Burkhard Kroger, Limburgerhof (DE); Hartwig Schroder, Nussloch (DE); Oskar Zelder, Speyer (DE); Gregor Haberhauer, Limburgerhof (DE)	Jul. 9, 1999 (DE).....	19932229.5
		Jul. 9, 1999 (DE).....	19932914.1
		Jul. 9, 1999 (DE).....	19933006.9
		Aug. 27, 1999 (DE).....	19940765.7
		Aug. 27, 1999 (DE).....	19940768.1
		Aug. 27, 1999 (DE).....	19940831.9
		Aug. 27, 1999 (DE).....	19940832.7
		Aug. 31, 1999 (DE).....	19941385.1
		Aug. 31, 1999 (DE).....	19941396.7
		Sep. 3, 1999 (DE).....	19942087.4

Correspondence Address:
LAHIVE & COCKFIELD, LLP.
28 STATE STREET
BOSTON, MA 02109 (US)

(73) Assignee: BASF Aktiengesellschaft, Ludwigshafen (DE)

Publication Classification

(21) Appl. No.: 10/721,922

(51) Int. Cl.⁷ C12Q 1/68; C07H 21/04;
C12P 13/04; C12N 9/10; C12N 1/21;
C12N 15/74

(22) Filed: Nov. 24, 2003

(52) U.S. Cl. 435/106; 435/6; 435/69.1;
435/193; 435/252.3; 435/320.1;
536/23.2**Related U.S. Application Data**

(63) Continuation of application No. 09/603,124, filed on Jun. 23, 2000, now abandoned.

(57) ABSTRACT

(60) Provisional application No. 60/141,031, filed on Jun. 25, 1999. Provisional application No. 60/143,694, filed on Jul. 14, 1999. Provisional application No. 60/151,778, filed on Aug. 31, 1999.

(30) Foreign Application Priority Data

Jul. 8, 1999 (DE).....	19931418.7
Jul. 9, 1999 (DE).....	19932124.8
Jul. 9, 1999 (DE).....	19932126.4
Jul. 9, 1999 (DE).....	19932127.2

Isolated nucleic acid molecules, designated HA nucleic acid molecules, which encode novel HA proteins from *Corynebacterium glutamicum* are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing HA nucleic acid molecules, and host cells into which the expression vectors have been introduced. The invention still further provides isolated HA proteins, mutated HA proteins, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from *C. glutamicum* based on genetic engineering of HA genes in this organism.

CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN HOMEOSTASIS AND ADAPTATION

RELATED APPLICATIONS

[0001] This application claims priority to prior filed U.S. Provisional Patent Application Ser. No. 60/141,031, filed Jun. 25, 1999, U.S. Provisional Patent Application Ser. No. 60/143,694, filed Jul. 14, 2000, and U.S. Provisional Patent Application Ser. No. 60/151,778, filed Aug. 31, 1999. This application also claims priority to German Application No. 19931418.7, filed Jul. 8, 1999, German Application No. 19932124.8, filed Jul. 9, 1999, German Application No. 19932126.4, filed Jul. 9, 1999, German Application No. 19932127.2, filed Jul. 9, 1999, German Application No. 19932133.7, filed Jul. 9, 1999, German Application No. 19932207.4, filed Jul. 9, 1999, German Application No. 19932208.2, filed Jul. 9, 1999, German Application No. 19932225.2, filed Jul. 9, 1999, German Application No. 19932229.5, filed Jul. 9, 1999, German Application No. 19932914.1, filed Jul. 9, 1999, German Application No. 19933006.9, filed Jul. 9, 1999, German Application No. 19940765.7, filed Aug. 27, 1999, German Application No. 19940768.1, filed Aug. 27, 1999, German Application No. 19940831.9, filed Aug. 27, 1999, German Application No. 19940832.7, filed Aug. 27, 1999, German Application No. 19941385.1, filed Aug. 31, 1999, German Application No. 19941396.7, filed Aug. 31, 1999, and German Application No. 19942087.4, filed Sep. 3, 1999. The entire contents of all of the aforementioned applications are hereby expressly incorporated herein by this reference.

BACKGROUND OF THE INVENTION

[0002] Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through the large-scale culture of bacteria developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

SUMMARY OF THE INVENTION

[0003] The invention provides novel bacterial nucleic acid molecules which have a variety of uses. These uses include the identification of microorganisms which can be used to produce fine chemicals, the modulation of fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or related bacteria, as reference points for mapping the *C. glutamicum* genome, and as markers for transformation. These novel nucleic acid molecules encode proteins, referred to herein as homeostasis and adaptation (HA) proteins.

[0004] *C. glutamicum* is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The HA nucleic acid molecules of the invention, therefore, can be used to identify microorganisms which can be used to produce fine chemicals, e.g., by fermentation processes. Modulation of the expression of the HA nucleic acids of the invention, or modification of the sequence of the HA nucleic acid molecules of the invention, can be used to modulate the production of one or more fine chemicals from a microorganism (e.g., to improve the yield or production of one or more fine chemicals from a *Corynebacterium* or *Brevibacterium* species).

[0005] The HA nucleic acids of the invention may also be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof, or to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to species pathogenic in humans, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria); the detection of such organisms is of significant clinical relevance.

[0006] The HA nucleic acid molecules of the invention may also serve as reference points for mapping of the *C. glutamicum* genome, or of genomes of related organisms. Similarly, these molecules, or variants or portions thereof, may serve as markers for genetically engineered *Corynebacterium* or *Brevibacterium* species.

[0007] The HA proteins encoded by the novel nucleic acid molecules of the invention are capable of, for example, performing a function involved in the maintenance of homeostasis in *C. glutamicum*, or in the ability of this microorganism to adapt to different environmental conditions. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey et al., U.S. Pat. No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (e.g., *lactofermentum*) (Yoshihama et al., *J. Bacteriol.* 162: 591-597 (1985); Katsumata et al., *J. Bacteriol.* 159: 306-311 (1984); and Santamaria et al., *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

[0008] There are a number of mechanisms by which the alteration of an HA protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by engineering

enzymes which modify or degrade aromatic or aliphatic compounds such that these enzymes are increased or decreased in activity or number, it may be possible to modulate the production of one or more fine chemicals which are the modification or degradation products of these compounds. Similarly, enzymes involved in the metabolism of inorganic compounds provide key molecules (e.g., phosphorous, sulfur, and nitrogen molecules) for the biosynthesis of such fine chemicals as amino acids, vitamins, and nucleic acids. By altering the activity or number of these enzymes in *C. glutamicum*, it may be possible to increase the conversion of these inorganic compounds (or to use alternate inorganic compounds) to thus permit improved rates of incorporation of inorganic atoms into these fine chemicals. Genetic engineering of *C. glutamicum* enzymes involved in general cellular processes may also directly improve fine chemical production, since many of these enzymes directly modify fine chemicals (e.g., amino acids) or the enzymes which are involved in fine chemical synthesis or secretion. Modulation of the activity or number of cellular proteases may also have a direct effect on fine chemical production, since many proteases may degrade fine chemicals or enzymes involved in fine chemical production or breakdown.

[0009] Further, the aforementioned enzymes which participate in aromatic/aliphatic compound modification or degradation, general biocatalysis, inorganic compound metabolism or proteolysis are each themselves fine chemicals, desirable for their activity in various in vitro industrial applications. By altering the number of copies of the gene for one or more of these enzymes in *C. glutamicum* it may be possible to increase the number of these proteins produced by the cell, thereby increasing the potential yield or efficiency of production of these proteins from large-scale *C. glutamicum* or related bacterial cultures.

[0010] The alteration of an HA protein of the invention may also indirectly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by modulating the activity and/or number of those proteins involved in the construction or rearrangement of the cell wall, it may be possible to modify the structure of the cell wall itself such that the cell is able to better withstand the mechanical and other stresses present during large-scale fermentative culture. Also, large-scale growth of *C. glutamicum* requires significant cell wall production. Modulation of the activity or number of cell wall biosynthetic or degradative enzymes may allow more rapid rates of cell wall biosynthesis, which in turn may permit increased growth rates of this microorganism in culture and thereby increase the number of cells producing the desired fine chemical.

[0011] By modifying the HA enzymes of the invention, one may also indirectly impact the yield, production, or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, many of the general enzymes in *C. glutamicum* may have a significant impact on global cellular processes (e.g., regulatory processes) which in turn have a significant effect on fine chemical metabolism. Similarly, proteases, enzymes which modify or degrade possibly toxic aromatic or aliphatic compounds, and enzymes which promote the metabolism of inorganic compounds all serve to increase the viability of *C. glutamicum*. The proteases aid in the selective removal of misfolded or misregulated proteins,

such as those that might occur under the relatively stressful environmental conditions encountered during large-scale fermentor culture. By altering these proteins, it may be possible to further enhance this activity and to improve the viability of *C. glutamicum* in culture. The aromatic/aliphatic modification or degradation proteins not only serve to detoxify these waste compounds (which may be encountered as impurities in culture medium or as waste products from cells themselves), but also to permit the cells to utilize alternate carbon sources if the optimal carbon source is limiting in the culture. By increasing their number and/or activity, the survival of *C. glutamicum* cells in culture may be enhanced. The inorganic metabolism proteins of the invention supply the cell with inorganic molecules required for all protein and nucleotide (among others) synthesis, and thus are critical for the overall viability of the cell. An increase in the number of viable cells producing one or more desired fine chemicals in large-scale culture should result in a concomitant increase in the yield, production, and/or efficiency of production of the fine chemical in the culture.

[0012] The invention provides novel nucleic acid molecules which encode proteins, referred to herein as HA proteins, which are capable of, for example, performing a function involved in the maintenance of homeostasis in *C. glutamicum*, or of participating in the ability of this microorganism to adapt to different environmental conditions. Nucleic acid molecules encoding an HA protein are referred to herein as HA nucleic acid molecules. In a preferred embodiment, an HA protein participates in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or possesses a *C. glutamicum* enzymatic or proteolytic activity. Examples of such proteins include those encoded by the genes set forth in Table 1.

[0013] Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (e.g., cDNAs, DNAs, or RNAs) comprising a nucleotide sequence encoding an HA protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of HA-encoding nucleic acids (e.g., DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred HA proteins of the present invention also preferably possess at least one of the HA activities described herein.

[0014] In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, e.g., sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion

thereof maintains an HA activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (e.g., an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

[0015] In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (e.g., an HA fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

[0016] In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* HA protein, or a biologically active portion thereof.

[0017] Another aspect of the invention pertains to vectors, e.g., recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an HA protein by culturing the host cell in a suitable medium. The HA protein can be then isolated from the medium or the host cell.

[0018] Yet another aspect of the invention pertains to a genetically altered microorganism in which an HA gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated HA sequence as a transgene. In another embodiment, an endogenous HA gene within the genome of the microorganism has been altered, e.g., functionally disrupted, by homologous recombination with an altered HA gene. In another embodiment, an endogenous or introduced HA gene in a microorganism has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional HA protein. In still another embodiment, one or more of the regulatory regions (e.g., a promoter, repressor, or inducer) of an HA gene in a microorganism has been altered (e.g., by deletion, truncation, inversion, or point mutation) such that the expression of the HA gene is modulated. In a

preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

[0019] In another aspect, the invention provides a method of identifying the presence or activity of *Corynebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (e.g., the sequences set forth in Appendix A or Appendix B) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject.

[0020] Still another aspect of the invention pertains to an isolated HA protein or a portion, e.g., a biologically active portion, thereof. In a preferred embodiment, the isolated HA protein or portion thereof can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions. In another preferred embodiment, the isolated HA protein or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions.

[0021] The invention also provides an isolated preparation of an HA protein. In preferred embodiments, the HA protein comprises an amino acid sequence of Appendix B. In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated HA protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions, or has one or more of the activities set forth in Table 1.

[0022] Alternatively, the isolated HA protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous, to a nucleotide sequence of Appendix B. It is also preferred that the preferred forms of HA proteins also have one or more of the HA bioactivities described herein.

[0023] The HA polypeptide, or a biologically active portion thereof, can be operatively linked to a non-HA polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the HA protein alone. In other preferred embodiments, this

fusion protein participates in the maintenance of homeostasis in *C. glutamicum*, or performs a function involved in the adaptation of this microorganism to different environmental conditions. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

[0024] In another aspect, the invention provides methods for screening molecules which modulate the activity of an HA protein, either by interacting with the protein itself or a substrate or binding partner of the HA protein, or by modulating the transcription or translation of an HA nucleic acid molecule of the invention.

[0025] Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an HA nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an HA nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

[0026] Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates HA protein activity or HA nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* processes involved in cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activities. The agent which modulates HA protein activity can be an agent which stimulates HA protein activity or HA nucleic acid expression. Examples of agents which stimulate HA protein activity or HA nucleic acid expression include small molecules, active HA proteins, and nucleic acids encoding HA proteins that have been introduced into the cell. Examples of agents which inhibit HA activity or expression include small molecules and antisense HA nucleic acid molecules.

[0027] Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant HA gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

DETAILED DESCRIPTION OF THE INVENTION

[0028] The present invention provides HA nucleic acid and protein molecules which are involved in *C. glutamicum*

cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (e.g., where overexpression or optimization of activity of a protein involved in the production of a fine chemical (e.g., an enzyme) has a direct impact on the yield, production, and/or efficiency of production of a fine chemical from the modified *C. glutamicum*), or an indirect impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound (e.g., where modulation of the activity or number of copies of a *C. glutamicum* aromatic or aliphatic modification or degradation protein results in an increase in the viability of *C. glutamicum* cells, which in turn permits increased production in a large-scale culture setting). Aspects of the invention are further explicated below.

[0029] I. Fine Chemicals

[0030] The term ‘fine chemical’ is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described e.g. in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (e.g., arachidonic acid), diols (e.g., propane diol, and butane diol), carbohydrates (e.g., hyaluronic acid and trehalose), aromatic compounds (e.g., aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann’s Encyclopedia of Industrial Chemistry, vol. A27, “Vitamins”, p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A. S., Niki, E. & Packer, L. (1995) “Nutrition, Lipids, Health, and Disease” Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research—Asia, held Sep. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, polyketides (Cane et al. (1998) *Science* 282: 63-68), and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

[0031] A. Amino Acid Metabolism and Uses

[0032] Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term “amino acid” is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ullmann’s Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L-optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino

acids have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3rd edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosyntheses, are readily converted by simple biosynthetic pathways to the remaining 11 'non-essential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

[0033] Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino acids—technical production and use, p. 466-502 in Rehm et al. (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

[0034] The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H. E. (1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of α -ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain β -carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is

formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

[0035] Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3rd ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3rd ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

[0036] B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

[0037] Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although they are readily synthesized by other organisms such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (e.g., polyunsaturated fatty acids).

[0038] The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A. S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations

in Malaysia, and the Society for Free Radical Research—Asia, held Sep. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, Ill. X, 374 S.

[0039] Thiamin (vitamin B₁) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B₂) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B₆' (e.g., pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of β-alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to β-alanine and for the condensation to pantothenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of pantothenate, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B₅), pantetheine (and its derivatives) and coenzyme A.

[0040] Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the α-ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

[0041] Corrinoids (such as the cobalamines and particularly vitamin B₁₂) and porphyrines belong to a group of chemicals characterized by a tetrapyrrole ring system. The biosynthesis of vitamin B₁₂ is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

[0042] The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B₆, pantothenate, and biotin. Only Vitamin B₁₂ is produced solely by fermentation, due to the complexity of its synthesis. In vitro methodologies require significant inputs of materials and time, often at great cost.

[0043] C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

[0044] Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (i.e., AMP) or as coenzymes (i.e., FAD and NAD).

[0045] Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (e.g. Christopherson, R. I. and Lyons, S. D. (1990) "Potent inhibitors of de novo pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J. L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (e.g., thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (e.g., ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (e.g., IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) *Nucleotides and Related Compounds in Biotechnology* vol. 6, Rehm et al., eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

[0046] The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J. E. (1992) "de novo purine nucleotide biosynthesis", in: *Progress in Nucleic Acid Research and Molecular Biology*, vol. 42, Academic Press; p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which

the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy-forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

[0047] D. Trehalose Metabolism and Uses

[0048] Trehalose consists of two glucose molecules, bound in α , α -1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto et al., (1998) U.S. Pat. No. 5,759,610; Singer, M. A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C. L. A. and Panek, A. D. (1996) *Biotech. Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) *J. Japan* 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

[0049] II. Maintenance of Homeostasis in *C. glutamicum* and Environmental Adaptation

[0050] The metabolic and other biochemical processes by which cells function are sensitive to environmental conditions such as temperature, pressure, solute concentration, and availability of oxygen. When one or more such environmental condition is perturbed or altered in a fashion that is incompatible with the normal functioning of these cellular processes, the cell must act to maintain an intracellular environment which will permit them to occur despite the hostile extracellular environment. Gram positive bacterial cells, such as *C. glutamicum* cells, have a number of mechanisms by which internal homeostasis may be maintained despite unfavorable extracellular conditions. These include a cell wall, proteins which are able to degrade possibly toxic aromatic and aliphatic compounds, mechanisms of proteolysis whereby misfolded or misregulated proteins may be rapidly destroyed, and catalysts which permit intracellular reactions to occur which would not normally take place under the conditions optimal for bacterial growth.

[0051] Aside from merely surviving in a hostile environment, bacterial cells (e.g. *C. glutamicum* cells) are also frequently able to adapt such that they are able to take advantage of such conditions. For example, cells in an environment lacking desired carbon sources may be able to adapt to growth on a less-suitable carbon source. Also, cells may be able to utilize less desirable inorganic compounds when the commonly utilized ones are unavailable. *C. glutamicum* cells possess a number of genes which permit them to adapt to utilize inorganic and organic molecules which they would normally not encounter under optimal growth conditions as nutrients and precursors for metabolism. Aspects of cellular processes involved in homeostasis and adaptation are further explicated below.

[0052] A. Modification and Degradation of Aromatic and Aliphatic Compounds

[0053] Bacterial cells are routinely exposed to a variety of aromatic and aliphatic compounds in nature. Aromatic compounds are organic molecules having a cyclic ring structure, while aliphatic compounds are organic molecules having open chain structures rather than ring structures. Such compounds may arise as by-products of industrial processes (e.g., benzene or toluene), but may also be produced by certain microorganisms (e.g., alcohols). Many of these compounds are toxic to cells, particularly the aromatic compounds, which are highly reactive due to the high-energy ring structure. Thus, certain bacteria have developed mechanisms by which they are able to modify or degrade these compounds such that they are no longer hazardous to the cell. Cells may possess enzymes that are able to, for example, hydroxylate, isomerize, or methylate aromatic or aliphatic compounds such that they are either rendered less toxic, or such that the modified form is able to be processed by standard cellular waste and degradation pathways. Also, cells may possess enzymes which are able to specifically degrade one or more such potentially hazardous substance, thereby protecting the cell. Principles and examples of these types of modification and degradation processes in bacteria are described in several publications, e.g., Sahm, H. (1999) "Prokaryotes in Industrial Production" in Lengeler, J. W. et al., eds. *Biology of the Prokaryotes*, Thieme Verlag: Stuttgart; and Schlegel, H. G. (1992) *Allgemeine Mikrobiologie*, Thieme: Stuttgart).

[0054] Aside from simply inactivating hazardous aromatic or aliphatic compounds, many bacteria have evolved to be able to utilize these compounds as carbon sources for continued metabolism when the preferred carbon sources of the cell are not available. For example, *Pseudomonas* strains able to utilize toluene, benzene, and 1,10-dichlorodecane as carbon sources are known (Chang, B. V. et al. (1997) *Chemosphere* 35(12): 2807-2815; Wischnak, C. et al. (1998) *Appl. Environ. Microbiol.* 64(9): 3507-3511; Churchill, S. A. et al. (1999) *Appl. Environ. Microbiol.* 65(2): 549-552). There are similar examples from many other bacterial species which are known in the art.

[0055] The ability of certain bacteria to modify or degrade aromatic and aliphatic compounds has begun to be exploited. Petroleum is a complex mixture of chemicals which includes aliphatic molecules and aromatic compounds. By applying bacteria having the ability to degrade or modify these toxic compounds to an oil spill, for example, it is possible to eliminate much of the environmental damage with high efficiency and low cost (see, for example, Smith, M. R. (1990) "The biodegradation of aromatic hydrocarbons by bacteria" *Biodegradation* 1(2-3): 191-206; and Suyama, T. et al. (1998) "Bacterial isolates degrading aliphatic polycarbonates," *FEMS Microbiol. Lett.* 161(2): 255-261).

[0056] B. Metabolism of Inorganic Compounds

[0057] Cells (e.g., bacterial cells) contain large quantities of different molecules, such as water, inorganic ions, and organic substances (e.g., proteins, sugars, and other macromolecules). The bulk of the mass of a typical cell consists of only 4 types of atoms: carbon, oxygen, hydrogen, and nitrogen. Although they represent a smaller percentage of the content of a cell, inorganic substances are equally as important to the proper functioning of the cell. Such mol-

ecules include phosphorous, sulfur, calcium, magnesium, iron, zinc, manganese, copper, molybdenum, tungsten, and cobalt. Many of these compounds are critical for the construction of important molecules, such as nucleotides (phosphorous) and amino acids (nitrogen and sulfur). Others of these inorganic ions serve as cofactors for enzymic reactions or contribute to osmotic pressure. All such molecules must be taken up by the bacterium from the surrounding environment.

[0058] For each of these inorganic compounds it is desirable for the bacterium to take up the form which can be most readily used by the standard metabolic machinery of the cell. However, the bacterium may encounter environments in which these preferred forms are not readily available. In order to survive under these circumstances, it is important for bacteria to have additional biochemical mechanisms which are able to convert less metabolically active but readily available forms of these inorganic compounds to ones which may be used in cellular metabolism. Bacteria frequently possess a number of genes encoding enzymes for this purpose, which are not expressed unless the desired inorganic species are not available. Thus, these genes for the metabolism of various inorganic compounds serve as another tool which bacteria may use to adapt to suboptimal environmental conditions.

[0059] After carbon, the most important element in the cell is nitrogen. A typical bacterial cell contains between 12-15% nitrogen. It is a constituent of amino acids and nucleotides, as well as many other important molecules in the cell. Further, nitrogen may serve as a substitute for oxygen as a terminal electron acceptor in energy metabolism. Good sources of nitrogen include many organic and inorganic compounds, such ammonia gas or ammonia salts (e.g., NH₄Cl, (NH₄)₂SO₄, or NH₄OH), nitrates, urea, amino acids, or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract, etc. Ammonia nitrogen is fixed by the action of particular enzymes: glutamate dehydrogenase, glutamine synthase, and glutamine-2-oxoglutarate aminotransferase. The transfer of amino-nitrogen from one organic molecule to another is accomplished by the aminotransferases, a class of enzymes which transfer one amino group from an alpha-amino acid to an alpha-keto acid. Nitrate may be reduced via nitrate reductase, nitrite reductase, and further redox enzymes until it is converted to molecular nitrogen or ammonia, which may be readily utilized by the cell in standard metabolic pathways.

[0060] Phosphorous is typically found intracellularly in both organic and inorganic forms, and may be taken up by the cell in either of these forms as well, though most microorganisms preferentially take up inorganic phosphate. The conversion of organic phosphate to a form which the cell can utilize requires the action of phosphatases (e.g., phytases, which hydrolyze phytate-yielding phosphate and inositol derivatives). Phosphate is a key element in the synthesis of nucleic acids, and also has a significant role in cellular energy metabolism (e.g., in the synthesis of ATP, ADP, and AMP).

[0061] Sulfur is a requirement for the synthesis of amino acids (e.g., methionine and cysteine), vitamins (e.g., thiamine, biotin, and lipoic acid) and iron sulfur proteins. Bacteria obtain sulfur primarily from inorganic sulfate,

though thiosulfate, sulfite, and sulfide are also commonly utilized. Under conditions where these compounds may not be readily available, many bacteria express genes which enable them to utilize sulfonate compounds such as 2-amino-sulfonate (taurine) (Kertesz, M. A. (1993) "Proteins induced by sulfate limitation in *Escherichia coli*, *Pseudomonas putida*, or *Staphylococcus aureus*." *J. Bacteriol.* 175: 1187-1190).

[0062] Other inorganic atoms, e.g., metal or calcium ions, are also critical for the viability of cells. Iron, for example, plays a key role in redox reactions and is a cofactor of iron-sulfur proteins, heme proteins, and cytochromes. The uptake of iron into bacterial cells may be accomplished by the action of siderophores, chelating agents which bind extracellular iron ions and translocate them to the interior of the cell. For reference on the metabolism of iron and other inorganic compounds, see: Lengeler et al. (1999) Biology of Prokaryotes, Thieme Verlag: Stuttgart; Neidhardt, F. C. et al., eds. *Escherichia coli* and *Salmonella*. ASM Press: Washington, D.C.; Sonenshein, A. L. et al., eds. (1997) *Bacillus subtilis* and Other Gram-Positive Bacteria, ASM Press: Washington, D.C.; Voet, D. and Voet, J. G. (1992) Biochemistry, VCH: Weinheim; Brock, T. D. and Madigan, M. T. (1991) Biology of Microorganisms, 6th ed. Prentice Hall: Englewood Cliffs, p. 267-269; Rhodes, P. M. and Stanbury, P. F. Applied Microbial Physiology—A Practical Approach, Oxford Univ. Press: Oxford.

[0063] C. Enzymes and Proteolysis

[0064] The intracellular conditions for which bacteria such as *C. glutamicum* are optimized are frequently not conditions under which many biochemical reactions would normally take place. In order to make such reactions proceed under physiological conditions, cells utilize enzymes. Enzymes are proteinaceous biological catalysts, spatially orienting reacting molecules or providing a specialized environment such that the energy barrier to a biochemical reaction is lowered. Different enzymes catalyze different reactions, and each enzyme may be the subject of transcriptional, translational, or posttranslational regulation such that the reaction will only take place under appropriate conditions and at specified times. Enzymes may contribute to the degradation (e.g., the proteases), synthesis (e.g., the synthases), or modification (e.g., transferases or isomerases) of compounds, all of which enable the production of necessary compounds within the cell. This, in turn, contributes to the maintenance of cellular homeostasis.

[0065] However, the fact that enzymes are optimized for activity under the physiological conditions at which the bacterium is most viable means that when environmental conditions are perturbed, there is a significant possibility that enzyme activity will also be perturbed. For example, changes in temperature may result in aberrantly folded proteins, and the same is true for changes of pH—protein folding is largely dependent on electrostatic and hydrophobic interactions of amino acids within the polypeptide chain, so any alteration to the charges on individual amino acids (as might be brought about by a change in cellular pH) may have a profound effect on the ability of the protein to correctly fold. Changes in temperature effectively change the amount of kinetic energy that the polypeptide molecule possesses, which affects the ability of the polypeptide to settle into a correctly folded, energetically stable configu-

ration. Misfolded proteins may be harmful to the cell for two reasons. First, the aberrantly folded protein may have a similarly aberrant activity, or no activity whatsoever. Second, misfolded proteins may lack the conformational regions necessary for proper regulation by other cellular systems and thus may continue to be active but in an uncontrolled fashion.

[0066] The cell has a mechanism by which misfolded enzymes and regulatory proteins may be rapidly destroyed before any damage occurs to the cell: proteolysis. Proteins such as those of the *Ia/Ion* family and those of the *Clp* family specifically recognize and degrade misfolded proteins (see, e.g., Sherman, M. Y., Goldberg, A. L. (1999) *EXS* 77: 57-78 and references therein and Porankiewicz J. (1999) *Molec. Microbiol.* 32(3): 449-58, and references therein; Neidhardt, F. C., et al. (1996) *E. coli* and *Salmonella*, ASM Press: Washington, D.C. and references therein; and Pritchard, G. G., and Coolbear, T. (1993) *FEMS Microbiol. Rev.* 12(1-3): 179-206 and references therein). These enzymes bind to misfolded or unfolded proteins and degrade them in an ATP-dependent manner. Proteolysis thus serves as an important mechanism employed by the cell to prevent damage to normal cellular functions upon environmental changes, and it further permits cells to survive under conditions and in environments which would otherwise be toxic due to misregulated and/or aberrant enzyme or regulatory activity.

[0067] Proteolysis also has important functions in the cell under optimal environmental conditions. Within normal metabolic processes, proteases aid in the hydrolysis of peptide bonds, in the catabolism of complex molecules to provide necessary degradation products, and in protein modification. Secreted proteases play an important role in the catabolism of external nutrients even prior to the entry of these compounds into the cell. Further, proteolytic activity itself may serve regulatory functions; sporulation in *B. subtilis* and cell cycle progression in *Caulobacter* spp. are known to be regulated by key proteolytic events in each of these species (Gottesman, S. (1999) *Curr. Opin. Microbiol.* 2(2): 142-147). Thus, proteolytic processes are key for cellular survival under both suboptimal and optimal environmental conditions, and contribute to the overall maintenance of homeostasis in cells.

[0068] D. Cell Wall Production and Rearrangements

[0069] While the biochemical machinery of the cell may be able to readily adapt to different and possibly unfavorable environments, cells still require a general mechanism by which they may be protected from the environment. For many bacteria, the cell wall affords such protection, and also plays roles in adhesion, cell growth and division, and transport of desired solutes and waste materials.

[0070] In order to function, cells require intracellular concentrations of metabolites and other molecules that are substantially higher than those of the surrounding media. Since these metabolites are largely prevented from leaving the cell due to the presence of the hydrophobic membrane, the tendency of the system is for water molecules to enter the cell from the external medium such that the interior concentrations of solutes match the exterior concentrations. Water molecules are readily able to cross the cellular membrane, and this membrane is not able to withstand the resulting swelling and pressure, which may lead to osmotic lysis of the cell. The rigidity of the cell wall greatly improves

the ability of the cell to tolerate these pressures, and offers a further barrier to the unwanted diffusion of these metabolites and desired solutes from the cell. Similarly, the cell wall also serves to prevent unwanted material from entering the cell.

[0071] The cell wall also participates in a number of other cellular processes, such as adhesion and cell growth and division. Due to the fact that the cell wall completely surrounds the cell, any interaction of the cell with its surroundings must be mediated by the cell wall. Thus, the cell wall must participate in any adherence of the cell to other cells and to desired surfaces. Further, the cell cannot grow or divide without concomitant changes in the cell wall. Since the protection that the wall affords requires its presence during growth, morphogenesis and multiplication, one of the key steps in cell division is cell wall synthesis within the cell such that a new cell divides from the old. Thus, frequently cell wall biosynthesis is regulated in tandem with cell growth and cell division (see, e.g., Sonenshein, A. L. et al, eds. (1993) *Bacillus subtilis* and Other Gram-Positive Bacteria, ASM: Washington, D.C.).

[0072] The structure of the cell wall varies between gram-positive and gram-negative bacteria. However, in both types, the fundamental structural unit of the wall remains similar: an overlapping lattice of two polysaccharides, N-acetyl glucosamine (NAG) and N-acetyl muramic acid (NAM) which are cross-linked by amino acids (most commonly L-alanine, D-glutamate, diaminopimelic acid, and D-alanine), termed 'peptidoglycan'. The processes involved in the synthesis of the cell wall are known (see, e.g., Michal, G., ed. (1999) *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, Wiley: New York).

[0073] In gram-negative bacteria, the inner cellular membrane is coated by a single-layered peptidoglycan (approximately 10 nm thick), termed the murein-sacculus. This peptidoglycan structure is very rigid, and its structure determines the shape of the organism. The outer surface of the murein-sacculus is covered with an outer membrane, containing porins and other membrane proteins, phospholipids, and lipopolysaccharides. To maintain a tight association with the outer membrane, the gram-negative cell wall also has interspersed lipid molecules which serve to anchor it to the surrounding membrane.

[0074] In gram-positive bacteria, such as *Corynebacterium glutamicum*, the cytoplasmic membrane is covered by a multi-layered peptidoglycan, which ranges from 20-80 nm in thickness (see, e.g., Lengeler et al. (1999) *Biology of Prokaryotes* Thieme Verlag: Stuttgart, p. 913-918, p. 875-899, and p. 88-109 and references therein). The gram-positive cell wall also contains teichoic acid, a polymer of glycerol or ribitol linked through phosphate groups. Teichoic acid is also able to associate with amino acids, and forms covalent bonds with muramic acid. Also present in the cell wall may be lipoteichoic acids and teichuronic acids. If present, cellular surface structures such as flagella or capsules will be anchored in this layer as well.

[0075] III. Elements and Methods of the Invention

[0076] The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as HA nucleic acid and protein molecules, which participate in the maintenance of homeostasis in *C. glutamicum*, or which

perform a function involved in the adaptation of this microorganism to different environmental conditions. In one embodiment, the HA molecules participate in *C. glutamicum* cell wall biosynthesis or rearrangements, in the metabolism of inorganic compounds, in the modification or degradation of aromatic or aliphatic compounds, or have an enzymatic or proteolytic activity. In a preferred embodiment, the activity of the HA molecules of the present invention with regard to *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activity has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the HA molecules of the invention are modulated in activity, such that the *C. glutamicum* cellular processes in which the HA molecules participate (e.g., *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activity) are also altered in activity, resulting either directly or indirectly in a modulation of the yield, production, and/or efficiency of production of a desired fine chemical by *C. glutamicum*.

[0077] The language, "HA protein" or "HA polypeptide" includes proteins which participate in a number of cellular processes related to *C. glutamicum* homeostasis or the ability of *C. glutamicum* cells to adapt to unfavorable environmental conditions. For example, an HA protein may be involved in *C. glutamicum* cell wall biosynthesis or rearrangements, in the metabolism of inorganic compounds in *C. glutamicum*, in the modification or degradation of aromatic or aliphatic compounds in *C. glutamicum*, or have a *C. glutamicum* enzymatic or proteolytic activity. Examples of HA proteins include those encoded by the HA genes set forth in Table 1 and Appendix A. The terms "HA gene" or "HA nucleic acid sequence" include nucleic acid sequences encoding an HA protein, which consist of a coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of HA genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e., fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly

regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (e.g., the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound. The term "homeostasis" is art-recognized and includes all of the mechanisms utilized by a cell to maintain a constant intracellular environment despite the prevailing extracellular environmental conditions. A non-limiting example of such processes is the utilization of a cell wall to prevent osmotic lysis due to high intracellular solute concentrations. The term "adaptation" or "adaptation to an environmental condition" is art-recognized and includes mechanisms utilized by the cell to render the cell able to survive under nonpreferred environmental conditions (generally speaking, those environmental conditions in which one or more favored nutrients are absent, or in which an environmental condition such as temperature, pH, osmolarity, oxygen percentage and the like fall outside of the optimal survival range of the cell). Many cells, including *C. glutamicum* cells, possess genes encoding proteins which are expressed under such environmental conditions and which permit continued growth in such suboptimal conditions.

[0078] In another embodiment, the HA molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. There are a number of mechanisms by which the alteration of an HA protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by engineering enzymes which modify or degrade aromatic or aliphatic compounds such that these enzymes are increased or decreased in activity or number, it may be possible to modulate the production of one or more fine chemicals which are the modification or degradation products of these compounds. Similarly, enzymes involved in the metabolism of inorganic compounds provide key molecules (e.g. phosphorous, sulfur, and nitrogen molecules) for the biosynthesis of such fine chemicals as amino acids, vitamins, and nucleic acids. By altering the activity or number of these enzymes in *C. glutamicum*, it may be possible to increase the conversion of these inorganic compounds (or to use alternate inorganic compounds) to thus permit improved rates of incorporation of inorganic atoms into these fine chemicals. Genetic engineering of *C. glutamicum* enzymes involved in general cellular processes may also directly improve fine chemical production, since many of these enzymes directly modify fine chemicals (e.g., amino acids) or the enzymes which are involved in fine chemical synthesis or secretion. Modulation of the activity or number of cellular proteases may also have a direct effect on fine chemical production, since many proteases may degrade fine chemicals or enzymes involved in fine chemical production or breakdown.

[0079] Further, the aforementioned enzymes which participate in aromatic/aliphatic compound modification or degradation, general biocatalysis, inorganic compound metabolism or proteolysis are each themselves fine chemicals, desirable for their activity in various in vitro industrial applications. By altering the number of copies of the gene for one or more of these enzymes in *C. glutamicum* it may

be possible to increase the number of these proteins produced by the cell, thereby increasing the potential yield or efficiency of production of these proteins from large-scale *C. glutamicum* or related bacterial cultures.

[0080] The alteration of an HA protein of the invention may also indirectly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by modulating the activity and/or number of those proteins involved in the construction or rearrangement of the cell wall, it may be possible to modify the structure of the cell wall itself such that the cell is able to better withstand the mechanical and other stresses present during large-scale fermentative culture. Also, large-scale growth of *C. glutamicum* requires significant cell wall production. Modulation of the activity or number of cell wall biosynthetic or degradative enzymes may allow more rapid rates of cell wall biosynthesis, which in turn may permit increased growth rates of this microorganism in culture and thereby increase the number of cells producing the desired fine chemical.

[0081] By modifying the HA enzymes of the invention, one may also indirectly impact the yield, production, or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, many of the general enzymes in *C. glutamicum* may have a significant impact on global cellular processes (e.g., regulatory processes) which in turn have a significant effect on fine chemical metabolism. Similarly, proteases, enzymes which modify or degrade possibly toxic aromatic or aliphatic compounds, and enzymes which promote the metabolism of inorganic compounds all serve to increase the viability of *C. glutamicum*. The proteases aid in the selective removal of misfolded or misregulated proteins, such as those that might occur under the relatively stressful environmental conditions encountered during large-scale fermentor culture. By altering these proteins, it may be possible to further enhance this activity and to improve the viability of *C. glutamicum* in culture. The aromatic/aliphatic modification or degradation proteins not only serve to detoxify these waste compounds (which may be encountered as impurities in culture medium or as waste products from cells themselves), but also to permit the cells to utilize alternate carbon sources if the optimal carbon source is limiting in the culture. By increasing their number and/or activity, the survival of *C. glutamicum* cells in culture may be enhanced. The inorganic metabolism proteins of the invention supply the cell with inorganic molecules required for all protein and nucleotide (among others) synthesis, and thus are critical for the overall viability of the cell. An increase in the number of viable cells producing one or more desired fine chemicals in large-scale culture should result in a concomitant increase in the yield, production, and/or efficiency of production of the fine chemical in the culture.

[0082] The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* HA DNAs and the predicted amino acid sequences of the *C. glutamicum* HA proteins are shown in Appendices A and B, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins that participate in *C. glutamicum* cell wall biosynthesis or rearrangements,

metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity.

[0083] The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of Appendix B. As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

[0084] The HA protein or a biologically active portion or fragment thereof of the invention can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions, or have one or more of the activities set forth in Table 1.

[0085] Various aspects of the invention are described in further detail in the following subsections.

[0086] A. Isolated Nucleic Acid Molecules

[0087] One aspect of the invention pertains to isolated nucleic acid molecules that encode HA polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of HA-encoding nucleic acid (e.g., HA DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3' end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated HA nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g., a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a DNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

[0088] A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence

of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* HA DNA can be isolated from a *C. glutamicum* library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from normal endothelial cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. (1979) *Biochemistry* 18: 5294-5299) and DNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, Md.; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, Fla.). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an HA nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

[0089] In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Corynebacterium glutamicum* HA DNAs of the invention. This DNA comprises sequences encoding HA proteins (i.e., the "coding region", indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in Appendix A. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A.

[0090] For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying RXA, RXN, RXS, or RXC number having the designation "RXA", "RXN", "RXS", or "RXC" followed by 5 digits (i.e., RXA02702, RXN02707, RXS02560, and RXC00110). Each of these sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA, RXN, RXS, or RXC designation to eliminate confusion. The recitation "one of the sequences in Appendix A", then, refers to any of the sequences in Appendix A, which may be distinguished by their differing RXA, RXN, RXS, or RXC designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B are identified by the same RXA, RXN, RXS, or RXC designations as Appen-

dix A, such that they can be readily correlated. For example, the amino acid sequences in Appendix B designated RXA02702, RXN02707, RXS02560, and RXC00110 are translations of the coding regions of the nucleotide sequence of nucleic acid molecules RXA02702, RXN02707, RXS02560, and RXC00110, respectively, in Appendix A. Each of the RXA, RXN, RXS, and RXC nucleotide and amino acid sequences of the invention has also been assigned a SEQ ID NO, as indicated in Table 1.

[0091] Several of the genes of the invention are "F-designated genes". An F-designated gene includes those genes set forth in Table 1 which have an 'F' in front of the RXA, RXN, RXS, or RXC designation. For example, SEQ ID NO:1, designated, as indicated on Table 1, as "F RXA02702", is an F-designed gene, as are SEQ ID NOs: 9, 11, and 13 (designated on Table 1 as "F RXA02707", "F RXA02708", and "F RXA02709", respectively).

[0092] In one embodiment, the nucleic acid molecules of the present invention are not intended to include those compiled in Table 2. In the case of the dapD gene, a sequence for this gene was published in Wehrmann, A., et al. (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

[0093] In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

[0094] In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. Ranges and identity values intermediate to the above-recited ranges, (e.g., 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

[0095] Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment

which can be used as a probe or primer or a fragment encoding a biologically active portion of an HA protein. The nucleotide sequences determined from the cloning of the HA genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning HA homologues in other cell types and organisms, as well as HA homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone HA homologues. Probes based on the HA nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an HA protein, such as by measuring a level of an HA-encoding nucleic acid in a sample of cells, e.g., detecting HA mRNA levels or determining whether a genomic HA gene has been mutated or deleted.

[0096] In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. Proteins involved in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, "the function of an HA protein" contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of HA protein activities are set forth in Table 1.

[0097] In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most

preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

[0098] Portions of proteins encoded by the HA nucleic acid molecules of the invention are preferably biologically active portions of one of the HA proteins. As used herein, the term "biologically active portion of an HA protein" is intended to include a portion, e.g., a domain/motif, of an HA protein that can participate in the maintenance of homeostasis in *C. glutamicum*, or that can perform a function involved in the adaptation of this microorganism to different environmental conditions, or has an activity as set forth in Table 1. To determine whether an HA protein or a biologically active portion thereof can participate in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or has a *C. glutamicum* enzymatic or proteolytic activity, an assay of enzymatic activity may be performed. Such assay methods are well known to those of ordinary skill in the art, as detailed in Example 8 of the Exemplification.

[0099] Additional nucleic acid fragments encoding biologically active portions of an HA protein can be prepared by isolating a portion of one of the sequences in Appendix B, expressing the encoded portion of the HA protein or peptide (e.g., by recombinant expression in vitro) and assessing the activity of the encoded portion of the HA protein or peptide.

[0100] The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same HA protein as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

[0101] It will be understood by one of ordinary skill in the art that in one embodiment the sequences of the invention are not meant to include the sequences of the prior art, such as those Genbank sequences set forth in Tables 2 or 4 which were available prior to the present invention. In one embodiment, the invention includes nucleotide and amino acid sequences having a percent identity to a nucleotide or amino acid sequence of the invention which is greater than that of a sequence of the prior art (e.g., a Genbank sequence (or the protein encoded by such a sequence) set forth in Tables 2 or 4). For example, the invention includes a nucleotide sequence which is greater than and/or at least 36% identical to the nucleotide sequence designated RXA00009 (SEQ ID NO:85), a nucleotide sequence which is greater than and/or at least 40% identical to the nucleotide sequence designated RXA00277 (SEQ ID NO:91), and a nucleotide sequence which is greater than and/or at least 43% identical to the nucleotide sequence designated RXA00499 (SEQ ID NO:173). One of ordinary skill in the art would be able to calculate the lower threshold of percent identity for any given sequence of the invention by examining the GAP-calculated percent identity scores set forth in Table 4 for

each of the three top hits for the given sequence, and by subtracting the highest GAP-calculated percent identity from 100 percent. One of ordinary skill in the art will also appreciate that nucleic acid and amino acid sequences having percent identities greater than the lower threshold so calculated (e.g., at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more identical) are also encompassed by the invention.

[0102] In addition to the *C. glutamicum* HA nucleotide sequences shown in Appendix A, it will be appreciated by those of ordinary skill in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of HA proteins may exist within a population (e.g., the *C. glutamicum* population). Such genetic polymorphism in the HA gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an HA protein, preferably a *C. glutamicum* HA protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the HA gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in HA that are the result of natural variation and that do not alter the functional activity of HA proteins are intended to be within the scope of the invention.

[0103] Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* HA DNA of the invention can be isolated based on their homology to the *C. glutamicum* HA nucleic acid disclosed herein using the *C. glutamicum* DNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those of ordinary skill in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6x sodium chloride/sodium citrate (SSC) at about 45° C., followed by one or more washes in 0.2xSSC, 0.1% SDS at 50-65° C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucle-

otide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* HA protein.

[0104] In addition to naturally-occurring variants of the HA sequence that may exist in the population, one of ordinary skill in the art will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded HA protein, without altering the functional ability of the HA protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the HA proteins (Appendix B) without altering the activity of said HA protein, whereas an "essential" amino acid residue is required for HA protein activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having HA activity) may not be essential for activity and thus are likely to be amenable to alteration without altering HA activity.

[0105] Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding HA proteins that contain changes in amino acid residues that are not essential for HA activity. Such HA proteins differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the HA activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of participating in the maintenance of homeostasis in *C. glutamicum*, or of performing a function involved in the adaptation of this microorganism to different environmental conditions, or has one or more of the activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B, more preferably at least about 60-70% homologous to one of the sequences in Appendix B, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

[0106] To determine the percent homology of two amino acid sequences (e.g., one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences of Appendix B) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology=# of identical positions/total # of positions×100).

[0107] An isolated nucleic acid molecule encoding an HA protein homologous to a protein sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in an HA protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an HA coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an HA activity described herein to identify mutants that retain HA activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

[0108] In addition to the nucleic acid molecules encoding HA proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded DNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire HA coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an HA protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of SEQ ID NO. 1 (RXA02702) comprises nucleotides 1 to 1458). In another embodiment, the antisense nucleic acid molecule is antisense to a "non-coding region" of the coding strand of a nucleotide sequence encoding HA. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

[0109] Given the coding strand sequences encoding HA disclosed herein (e.g., the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The

antisense nucleic acid molecule can be complementary to the entire coding region of HA mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of HA mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of HA mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N-6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0110] The antisense nucleic acid molecules of the invention are typically administered to a cell or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an HA protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

[0111] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215: 327-330).

[0112] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334: 585-591)) can be used to catalytically cleave HA mRNA transcripts to thereby inhibit translation of HA mRNA. A ribozyme having specificity for an HA-encoding nucleic acid can be designed based upon the nucleotide sequence of an HA DNA molecule disclosed herein (i.e., SEQ ID NO. 3 (RXA02705) Appendix A). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an HA-encoding mRNA. See, e.g., Cech et al. U.S. Pat. No. 4,987,071 and Cech et al. U.S. Pat. No. 5,116,742. Alternatively, HA mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J. W. (1993) *Science* 261: 1411-1418.

[0113] Alternatively, HA gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an HA nucleotide sequence (e.g., an HA promoter and/or enhancers) to form triple helical structures that prevent transcription of an HA gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6): 569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660: 27-36; and Maher, L. J. (1992) *Bioassays* 14(12): 807-15.

[0114] B. Recombinant Expression Vectors and Host Cells

[0115] Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an HA protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used inter-

changeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

[0116] The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which are operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells. Preferred regulatory sequences are, for example, promoters such as cos-, tac-, trp-, tet-, trp-tet-, lpp-, lac-, lpp-lac-, lacI^q-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, arny, SPO2, λ -P_R- or λ P_L, which are used preferably in bacteria. Additional regulatory sequences are, for example, promoters from yeasts and fungi, such as ADC1, Mf_A, AC, P-60, CYC1, GAPDH, TEF, rp28, ADH, promoters from plants such as CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos or ubiquitin- or phaseolin-promoters. It is also possible to use artificial promoters. It will be appreciated by those of ordinary skill in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., HA proteins, mutant forms of HA proteins, fusion proteins, etc.).

[0117] The recombinant expression vectors of the invention can be designed for expression of HA proteins in prokaryotic or eukaryotic cells. For example, HA genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M. A. et al. (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C. A. M. J. J. et al. (1991) "Heterologous gene expression in filamentous fungi" in: More Gene Manipulations in Fungi, J. W. Bennet & L. L. Lasure, eds., p. 396-428; Academic Press: San Diego; and van den Hondel, C. A. M. J. J. & Punt, P. J. (1991) "Gene transfer systems and vector development for filamentous fungi", in: Applied Molecular Genetics of Fungi, Peberdy, J. F. et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian

cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[0118] Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

[0119] Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D. B. and Johnson, K. S. (1988) *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the HA protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant HA protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

[0120] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69: 301-315) pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHs1, pHs2, pPLc236, pMBL24, pLG200, pUR290, pIN-III113-B1, λgt11, pBdC1, and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990) 60-89; and Pouwels et al., eds. (1985) *Cloning Vectors*. Elsevier: New York ISBN 0 444 904018). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21 (DE3) or HMS174(DE3) from a resident % prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter. For transformation of other varieties of bacteria, appropriate vectors may be selected. For example, the plasmids pIJ101, pIJ364, pIJ702 and pIJ361 are known to be useful in transforming *Streptomyces*, while plasmids pUB110, pC194, or pBD214 are suited for transformation of *Bacillus* species. Several plasmids of use in the transfer of genetic information into *Corynebacterium* include

pHM1519, pBL1, pSA77, or pAJ667 (Pouwels et al., eds. (1985) *Cloning Vectors*. Elsevier: New York ISBN 0 444 904018).

[0121] One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada et al. (1992) *Nucleic Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[0122] In another embodiment, the HA protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYEpSec1 (Baldari, et al., (1987) *Embo J.* 6: 229-234), 2μ, pAG-1, Yep6, Yep13, pEMBLYe23, pMFA (Kurjan and Herskowitz, (1982) *Cell* 30: 933-943), pJRY88 (Schultz et al., (1987) *Gene* 54: 113-123), and pYES2 (Invitrogen Corporation, San Diego, Calif.). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C. A. M. J. J. & Punt, P. J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: *Applied Molecular Genetics of Fungi*, J. F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge, and Pouwels et al., eds. (1985) *Cloning Vectors*. Elsevier: New York (ISBN 0 444 904018).

[0123] Alternatively, the HA proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170: 31-39).

[0124] In another embodiment, the HA proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan, M. W. (1984) "Binary *Agrobacterium* vectors for plant transformation", *Nucl. Acid. Res.* 12: 8711-8721, and include pLGV23, pGHLac+, pBIN19, pAK2004, and pDH51 (Pouwels et al., eds. (1985) *Cloning Vectors*. Elsevier: New York ISBN 0 444 904018).

[0125] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329: 840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for

both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0126] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8: 729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33: 729-740; Queen and Baltimore (1983) *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86: 5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3: 537-546).

[0127] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to HA mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews—Trends in Genetics*, Vol. 1(1) (1986).

[0128] Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0129] A host cell can be any prokaryotic or eukaryotic cell. For example, an HA protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those of ordinary skill in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

[0130] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., linear DNA or RNA (e.g., a linearized vector or a gene construct alone without a vector) or nucleic acid in the form of a vector (e.g., a plasmid, phage, phasmid, phagemid, transposon or other DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

[0131] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an HA protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0132] To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an HA gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the HA gene. Preferably, this HA gene is a *Corynebacterium glutamicum* HA gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous HA gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous HA gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous HA protein). In the homologous recombination vector, the altered portion of the HA gene is flanked at its 5' and 3' ends by additional nucleic acid of the HA gene to allow for homologous recombination to occur between the

exogenous HA gene carried by the vector and an endogenous HA gene in a microorganism. The additional flanking HA nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K. R., and Capecchi, M. R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (e.g., by electroporation) and cells in which the introduced HA gene has homologously recombined with the endogenous HA gene are selected, using art-known techniques.

[0133] In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an HA gene on a vector placing it under control of the lac operon permits expression of the HA gene only in the presence of IPTG. Such regulatory systems are well known in the art.

[0134] In another embodiment, an endogenous HA gene in a host cell is disrupted (e.g., by homologous recombination or other genetic means known in the art) such that expression of its protein product does not occur. In another embodiment, an endogenous or introduced HA gene in a host cell has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional HA protein. In still another embodiment, one or more of the regulatory regions (e.g., a promoter, repressor, or inducer) of an HA gene in a microorganism has been altered (e.g., by deletion, truncation, inversion, or point mutation) such that the expression of the HA gene is modulated. One of ordinary skill in the art will appreciate that host cells containing more than one of the described HA gene and protein modifications may be readily produced using the methods of the invention, and are meant to be included in the present invention.

[0135] A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) an HA protein. Accordingly, the invention further provides methods for producing HA proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an HA protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered HA protein) in a suitable medium until HA protein is produced. In another embodiment, the method further comprises isolating HA proteins from the medium or the host cell.

[0136] C. Isolated HA Proteins

[0137] Another aspect of the invention pertains to isolated HA proteins, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of HA protein in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of HA protein having less than about 30% (by dry weight) of non-HA protein (also referred to herein as a "contaminating protein"), more preferably less

than about 20% of non-HA protein, still more preferably less than about 10% of non-HA protein, and most preferably less than about 5% non-HA protein. When the HA protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of HA protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of HA protein having less than about 30% (by dry weight) of chemical precursors or non-HA chemicals, more preferably less than about 20% chemical precursors or non-HA chemicals, still more preferably less than about 10% chemical precursors or non-HA chemicals, and most preferably less than about 5% chemical precursors or non-HA chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the HA protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* HA protein in a microorganism such as *C. glutamicum*.

[0138] An isolated HA protein or a portion thereof of the invention can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an HA protein of the invention has an amino acid sequence shown in Appendix B. In yet another preferred embodiment, the HA protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the HA protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to one of the nucleic acid sequences of Appendix A, or a portion thereof. Ranges and identity values intermediate to the above-recited values, (e.g., 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are

intended to be included. The preferred HA proteins of the present invention also preferably possess at least one of the HA activities described herein. For example, a preferred HA protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this micro-organism to different environmental conditions, or which has one or more of the activities set forth in Table 1.

[0139] In other embodiments, the HA protein is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the HA protein is a protein which comprises an amino acid sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the HA activities described herein. Ranges and identity values intermediate to the above-recited values, (e.g., 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

[0140] Biologically active portions of an HA protein include peptides comprising amino acid sequences derived from the amino acid sequence of an HA protein, e.g., the an amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an HA protein, which include fewer amino acids than a full length HA protein or the full length protein which is homologous to an HA protein, and exhibit at least one activity of an HA protein. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an HA protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an HA protein include one or more selected domains/motifs or portions thereof having biological activity.

[0141] HA proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the HA protein is expressed in the host cell. The HA protein can then be isolated from the cells by an appropriate purification scheme

using standard protein purification techniques. Alternative to recombinant expression, an HA protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native HA protein can be isolated from cells (e.g., endothelial cells), for example using an anti-HA antibody, which can be produced by standard techniques utilizing an HA protein or fragment thereof of this invention.

[0142] The invention also provides HA chimeric or fusion proteins. As used herein, an HA "chimeric protein" or "fusion protein" comprises an HA polypeptide operatively linked to a non-HA polypeptide. An "HA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an HA protein, whereas a "non-HA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the HA protein, e.g., a protein which is different from the HA protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the HA polypeptide and the non-HA polypeptide are fused in-frame to each other. The non-HA polypeptide can be fused to the N-terminus or C-terminus of the HA polypeptide. For example, in one embodiment the fusion protein is a GST-HA fusion protein in which the HA sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant HA proteins. In another embodiment, the fusion protein is an HA protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of an HA protein can be increased through use of a heterologous signal sequence.

[0143] Preferably, an HA chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An HA-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the HA protein.

[0144] Homologues of the HA protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the HA protein. As used herein, the term "homologue" refers to a variant form of the HA protein which acts as an agonist or antagonist of the activity of the HA protein. An agonist of the HA protein can retain substantially the same, or a subset, of the biological activities of the HA protein. An antagonist of the HA protein can inhibit one or more of the activities of

the naturally occurring form of the HA protein, by, for example, competitively binding to a downstream or upstream member of a biochemical cascade which includes the HA protein, by binding to a target molecule with which the HA protein interacts, such that no functional interaction is possible, or by binding directly to the HA protein and inhibiting its normal activity.

[0145] In an alternative embodiment, homologues of the HA protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the HA protein for HA protein agonist or antagonist activity. In one embodiment, a variegated library of HA variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of HA variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential HA sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of HA sequences therein. There are a variety of methods which can be used to produce libraries of potential HA homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential HA sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S. A. (1983) *Tetrahedron* 39: 3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53: 323; Itakura et al. (1984) *Science* 198: 1056; Ike et al. (1983) *Nucleic Acid Res.* 11: 477.

[0146] In addition, libraries of fragments of the HA protein coding can be used to generate a variegated population of HA fragments for screening and subsequent selection of homologues of an HA protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an HA coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the HA protein.

[0147] Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of HA homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble

mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify HA homologues (Arkin and Yourvan (1992) *PNAS* 89: 7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3): 327-331).

[0148] In another embodiment, cell based assays can be exploited to analyze a variegated HA library, using methods well known in the art.

[0149] D. Uses and Methods of the Invention

[0150] The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of HA protein regions required for function; modulation of an HA protein activity; modulation of the metabolism of one or more inorganic compounds; modulation of the modification or degradation of one or more aromatic or aliphatic compounds; modulation of cell wall synthesis or rearrangements; modulation of enzyme activity or proteolysis; and modulation of cellular production of a desired compound, such as a fine chemical.

[0151] The HA nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to pathogenic species, such as *Corynebacterium diphtheriae*. *Corynebacterium diphtheriae* is the causative agent of diphtheria, a rapidly developing, acute, febrile infection which involves both local and systemic pathology. In this disease, a local lesion develops in the upper respiratory tract and involves necrotic injury to epithelial cells; the bacilli secrete toxin which is disseminated through this lesion to distal susceptible tissues of the body. Degenerative changes brought about by the inhibition of protein synthesis in these tissues, which include heart, muscle, peripheral nerves, adrenals, kidneys, liver and spleen, result in the systemic pathology of the disease. Diphtheria continues to have high incidence in many parts of the world, including Africa, Asia, Eastern Europe and the independent states of the former Soviet Union. An ongoing epidemic of diphtheria in the latter two regions has resulted in at least 5,000 deaths since 1990.

[0152] In one embodiment, the invention provides a method of identifying the presence or activity of *Corynebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (e.g., the sequences set forth in Appendix A or Appendix B) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in

the subject. *C. glutamicum* and *C. diphtheriae* are related bacteria, and many of the nucleic acid and protein molecules in *C. glutamicum* are homologous to *C. diphtheriae* nucleic acid and protein molecules, and can therefore be used to detect *C. diphtheriae* in a subject.

[0153] The nucleic acid and protein molecules of the invention may also serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

[0154] The HA nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The processes involved in adaptation and the maintenance of homeostasis in which the molecules of the invention participate are utilized by a wide variety of species; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

[0155] Manipulation of the HA nucleic acid molecules of the invention may result in the production of HA proteins having functional differences from the wild-type HA proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

[0156] The invention provides methods for screening molecules which modulate the activity of an HA protein, either by interacting with the protein itself or a substrate or binding partner of the HA protein, or by modulating the transcription or translation of an HA nucleic acid molecule of the invention. In such methods, a microorganism expressing one or more HA proteins of the invention is contacted with one or more test compounds, and the effect of each test compound on the activity or level of expression of the HA protein is assessed.

[0157] The modulation of activity or number of HA proteins involved in cell wall biosynthesis or rearrangements may impact the production, yield, and/or efficiency of production of one or more fine chemicals from *C. glutamicum*

cells. For example, by altering the activity of these proteins, it may be possible to modulate the structure or thickness of the cell wall. The cell wall serves in large measure as a protective device against osmotic lysis and external sources of injury; by modifying the cell wall it may be possible to increase the ability of *C. glutamicum* to withstand the mechanical and shear force stresses encountered by this microorganism during large-scale fermentor culture. Further, each *C. glutamicum* cell is surrounded by a thick cell wall, and thus, a significant portion of the biomass present in large scale culture consists of cell wall. By increasing the rate at which the cell wall is synthesized or by activating cell wall synthesis (through genetic engineering of the HA cell wall proteins of the invention) it may be possible to improve the growth rate of the microorganism. Similarly, by decreasing the activity or number of proteins involved in the degradation of cell wall or by decreasing the repression of cell wall biosynthesis, an overall increase in cell wall production may be achieved. An increase in the number of viable *C. glutamicum* cells (as may be accomplished by any of the foregoing described protein alterations) should result in increased numbers of cells producing the desired fine chemical in large-scale fermentor culture, which should permit increased yields or efficiency of production of these compounds from the culture.

[0158] The modulation of activity or number of *C. glutamicum* HA proteins that participate in the modification or degradation of aromatic or aliphatic compounds may also have direct or indirect impacts on the production of one or more fine chemicals from these cells. Certain aromatic or aliphatic modification or degradation products are desirable fine chemicals (e.g., organic acids or modified aromatic and aliphatic compounds); thus, by modifying the enzymes which perform these modifications (e.g., hydroxylation, methylation, or isomerization) or degradation reactions, it may be possible to increase the yields of these desired compounds. Similarly, by decreasing the activity or number of proteins involved in pathways which further degrade the modified or breakdown products of the aforementioned reactions it may be possible to improve the yields of these fine chemicals from *C. glutamicum* cells in culture.

[0159] These aromatic and aliphatic modification and degradative enzymes are themselves fine chemicals. In purified form, these enzymes may be used to degrade aromatic and aliphatic compounds (e.g., toxic chemicals such as petroleum products), either for the bioremediation of polluted sites, for the engineered decomposition of wastes, or for the large-scale and economically feasible production of desired modified aromatic or aliphatic compounds or their breakdown products, some of which may be conveniently used as carbon or energy sources for other fine chemical-producing compounds in culture (see, e.g., Faber, K. (1995) *Biotransformations in Organic Chemistry*, Springer: Berlin and references therein; and Roberts, S. M., ed. (1992-1996) *Preparative Biotransformations*, Wiley: Chichester, and references therein). By genetically altering these proteins such that their regulation by other cellular mechanisms is lessened or abolished, it may be possible to increase the overall number or activity of these proteins, thereby improving not only the yield of these fine chemicals but also the activity of these harvested proteins.

[0160] The modification of these aromatic and aliphatic modifying and degradation enzymes may also have an

indirect effect on the production of one or more fine chemical. Many aromatic and aliphatic compounds (such as those that may be encountered as impurities in culture media or as waste products from cellular metabolism) are toxic to cells; by modifying and/or degrading these compounds such that they may be readily removed or destroyed, cellular viability should be increased. Further, these enzymes may modify or degrade these compounds in such a manner that the resulting products may enter the normal carbon metabolism pathways of the cell, thus rendering the cell able to use these compounds as alternate carbon or energy sources. In large-scale culture situations, when there may be limiting amounts of optimal carbon sources, these enzymes provide a method by which cells may continue to grow and divide using aromatic or aliphatic compounds as nutrients. In either case, the resulting increase in the number of *C. glutamicum* cells in the culture producing the desired fine chemical should in turn result in increased yields or efficiency of production of the fine chemical(s).

[0161] Modifications in activity or number of HA proteins involved in the metabolism of inorganic compounds may also directly or indirectly affect the production of one or more fine chemicals from *C. glutamicum* or related bacterial cultures. For example, many desirable fine chemicals, such as nucleic acids, amino acids, cofactors and vitamins (e.g., thiamine, biotin, and lipoic acid) cannot be synthesized without inorganic molecules such as phosphorous, nitrate, sulfate, and iron. The inorganic metabolism proteins of the invention permit the cell to obtain these molecules from a variety of inorganic compounds and to divert them into various fine chemical biosynthetic pathways. Therefore, by increasing the activity or number of enzymes involved in the metabolism of these inorganic compounds, it may be possible to increase the supply of these possibly limiting inorganic molecules, thereby directly increasing the production or efficiency of production of various fine chemicals from *C. glutamicum* cells containing such altered proteins. Modification of the activity or number of inorganic metabolism enzymes of the invention may also render *C. glutamicum* able to better utilize limited inorganic compound supplies, or to utilize nonoptimal inorganic compounds to synthesize amino acids, vitamins, cofactors, or nucleic acids, all of which are necessary for continued growth and replication of the cell. By improving the viability of these cells in large-scale culture, the number of *C. glutamicum* cells producing one or more fine chemicals in the culture may also be increased, in turn increasing the yields or efficiency of production of one or more fine chemicals.

[0162] *C. glutamicum* enzymes for general processes are themselves desirable fine chemicals. The specific properties of enzymes (i.e., regio- and stereospecificity, among others) make them useful catalysts for chemical reactions *in vitro*. Either whole *C. glutamicum* cells may be incubated with an appropriate substrate such that the desired product is produced by enzymes in the cell, or the desired enzymes may be overproduced and purified from *C. glutamicum* cultures (or those of a related bacterium) and subsequently utilized in *in vitro* reactions in an industrial setting (either in solution or immobilized on a suitable immobile phase). In either situation, the enzyme can either be a natural *C. glutamicum* protein, or it may be mutagenized to have an altered activity; typical industrial uses for such enzymes include as catalysts in the chemical industry (e.g., for synthetic organic chem-

istry) as food additives, as feed components, for fruit processing, for leather preparation, in detergents, in analysis and medicine, and in the textile industry (see, e.g. Yamada, H. (1993) "Microbial reactions for the production of useful organic compounds," *Chimica* 47: 5-10; Roberts, S. M. (1998) Preparative biotransformations: the employment of enzymes and whole-cells in synthetic chemistry," *J. Chem. Soc. Perkin Trans. 1*: 157-169; Zaks, A. and Dodds, D. R. (1997) "Application of biocatalysis and biotransformations to the synthesis of pharmaceuticals," *DDT* 2: 513-531; Roberts, S. M. and Williamson, N. M. (1997) "The use of enzymes for the preparation of biologically active natural products and analogues in optically active form," *Curr. Organ. Chemistry* 1: 1-20; Faber, K. (1995) Biotransformations in Organic Chemistry, Springer: Berlin; Roberts, S. M., ed. (1992-96) Preparative Biotransformations, Wiley: Chichester; Cheetham, P. S. J. (1995) "The applications of enzymes in industry" in: Handbook of Enzyme Biotechnology, 3rd ed., Wiseman, A., ed., Ellis: Horwood, p. 419-552; and Ullmann's Encyclopedia of Industrial Chemistry (1987), vol. A9, Enzymes, p. 390-457). Thus, by increasing the activity or number of these enzymes, it may be possible to also increase the ability of the cell to convert supplied substrates to desired products, or to overproduce these enzymes for increased yields in large-scale culture. Further, by mutagenizing these proteins it may be possible to remove feedback inhibition or other repressive cellular regulatory controls such that greater numbers of these enzymes may be produced and activated by the cell, thereby leading to greater yields, production, or efficiency of production of these fine chemical proteins from large-scale cultures. Further, manipulation of these enzymes may alter the activity of one or more *C. glutamicum* metabolic pathways, such as those for the biosynthesis or secretion of one or more fine chemicals.

[0163] Mutagenesis of the proteolytic enzymes of the invention such that they are altered in activity or number may also directly or indirectly affect the yield, production, and/or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, by increasing the activity or number of these proteins, it may be possible to increase the ability of the bacterium to survive in large-scale culture, due to an increased ability of the cell to rapidly degrade proteins misfolded in response to the high temperatures, nonoptimal pH, and other stresses encountered during fermentor culture. Increased numbers of cells in these cultures may result in increased yields or efficiency of production of one or more desired fine chemicals, due to the relatively larger number of cells producing these compounds in the culture. Also, *C. glutamicum* cells possess multiple cell-surface proteases which serve to break down external nutrients into molecules which may be more readily incorporated by the cells as carbon/energy sources or nutrients of other kinds. An increase in activity or number of these enzymes may improve this turnover and increase the levels of available nutrients, thereby improving cell growth or production. Thus, modifications of the proteases of the invention may indirectly impact *C. glutamicum* fine chemical production.

[0164] A more direct impact on fine chemical production in response to the modification of one or more of the proteases of the invention may occur when these proteases are involved in the production or degradation of a desired fine chemical. By decreasing the activity of a protease which

degrades a fine chemical or a protein involved in the synthesis of a fine chemical it may be possible to increase the levels of that fine chemical (due to the decreased degradation or increased synthesis of the compound). Similarly, by increasing the activity of a protease which degrades a compound to result in a fine chemical or a protein involved in the degradation of a fine chemical, a similar result should be achieved: increased levels of the desired fine chemical from *C. glutamicum* cells containing these engineered proteins.

[0165] The aforementioned mutagenesis strategies for HA proteins to result in increased yields of a fine chemical from *C. glutamicum* are not meant to be limiting; variations on these strategies will be readily apparent to one of ordinary skill in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated HA nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any product produced by *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

[0166] This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, published patent applications, Tables, Appendices, and the sequence listing cited throughout this application are hereby incorporated by reference.

[0167] Exemplification

EXAMPLE 1

Preparation of Total Genomic DNA of *Corynebacterium Glutamicum* ATCC 13032

[0168] A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30° C. with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-I (5% of the original volume of the culture—all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-I: 140.34 g/l sucrose, 2.46 g/l MgSO₄×7H₂O, 10 ml/l KH₂PO₄ solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l (NH₄)₂SO₄, 1 g/l NaCl, 2 g/l MgSO₄×7H₂O, 0.2 g/l CaCl₂, 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l FeSO₄×H₂O, 10 mg/l ZnSO₄×7H₂O, 3 mg/l MnCl₂×4H₂O, 30 mg/l H₂BO₃, 20 mg/l CoCl₂×6H₂O, 1 mg/l NiCl₂×6H₂O, 3 mg/l Na₂MoO₄×2H₂O, 500 mg/l complexing agent (EDTA or citric acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-panthothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37° C., the

cell wall was degraded and the resulting protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-I and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 µg/ml, the suspension is incubated for ca. 18 h at 37° C. The DNA was purified by extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20° C. and a 30 min centrifugation at 12,000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20 µg/ml RNaseA and dialysed at 4° C. against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20° C., the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

EXAMPLE 2

Construction of Genomic Libraries in *Escherichia Coli* of *Corynebacterium Glutamicum* ATCC13032

[0169] Using DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (see e.g., Sambrook, J. et al. (1989) "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F. M. et al. (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

[0170] Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J. G. (1979) *Proc. Natl. Acad. Sci. USA*, 75: 3737-3741); pACYC177 (Change & Cohen (1978) *J. Bacteriol.* 134: 1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T. J., Rosenthal A. and Waterson, R. H. (1987) *Gene* 53: 283-286. Gene libraries specifically for use in *C. glutamicum* may be constructed using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

EXAMPLE 3

DNA Sequencing and Computational Functional Analysis

[0171] Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see e.g., Fleischman, R. D. et al. (1995) "Whole-genome Random Sequencing and Assembly of *Haemophilus Influenzae* Rd.", *Science*, 269: 496-512). Sequencing primers with the following nucleotide sequences were used:

5' -GGAAACAGTATGACCATG-3'
or
5' -GTAAAACGACGGCCAGT-3'.

EXAMPLE 4

In Vivo Mutagenesis

[0172] In vivo mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W. D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those of ordinary skill in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34.

EXAMPLE 5

DNA Transfer Between *Escherichia Coli* and *Corynebacterium Glutamicum*

[0173] Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as e.g., pHM1519 or pBL1) which replicate autonomously (for review see, e.g., Martin, J. F. et al. (1987) *Biotechnology*, 5: 137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. et al. (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F. M. et al. (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E. L. (1987) "From Genes to Clones—Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli* and *C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see e.g., Yoshihama, M. et al. (1985) *J. Bacteriol.* 162: 591-597, Martin J. F. et al. (1987) *Biotechnology*, 5: 137-146 and Eikmanns, B. J. et al. (1991) *Gene*, 102: 93-98).

[0174] Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. et al. (1984) *J. Bacteriol.* 159:306-311), electroporation (Liebl, E. et al. (1989) *FEMS Microbiol. Letters*, 53: 399-303) and in cases where special vectors are used, also by conjugation (as described e.g. in Schäfer, A et

al. (1990) *J. Bacteriol.* 172: 1663-1666). It is also possible to transfer the shuttle vectors for *C. glutamicum* to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an *Mcr*-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) *J. Mol. Biol.* 166: 1-19).

[0175] Genes may be overexpressed in *C. glutamicum* strains using plasmids which comprise pCG1 (U.S. Pat. No. 4,617,267) or fragments thereof, and optionally the gene for kanamycin resistance from TN903 (Grindley, N. D. and Joyce, C. M. (1980) *Proc. Natl. Acad. Sci. USA* 77(12): 7176-7180). In addition, genes may be overexpressed in *C. glutamicum* strains using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

[0176] Aside from the use of replicative plasmids, gene overexpression can also be achieved by integration into the genome. Genomic integration in *C. glutamicum* or other *Corynebacterium* or *Brevibacterium* species may be accomplished by well-known methods, such as homologous recombination with genomic region(s), restriction endonuclease mediated integration (REMI) (see, e.g., DE Patent 19823834), or through the use of transposons. It is also possible to modulate the activity of a gene of interest by modifying the regulatory regions (e.g., a promoter, a repressor, and/or an enhancer) by sequence modification, insertion, or deletion using site-directed methods (such as homologous recombination) or methods based on random events (such as transposon mutagenesis or REMI). Nucleic acid sequences which function as transcriptional terminators may also be inserted 3' to the coding region of one or more genes of the invention; such terminators are well-known in the art and are described, for example, in Winnacker, E. L. (1987) *From Genes to Clones—Introduction to Gene Technology*, VCH: Weinheim.

EXAMPLE 6

Assessment of the Expression of the Mutant Protein

[0177] Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. (1988) *Current Protocols in Molecular Biology*, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by several methods, all well-known in the art, such as that described in Bormann, E. R. et al. (1992) *Mol. Microbiol.* 6: 317-326.

[0178] To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such

as a Western blot, may be employed (see, for example, Ausubel et al. (1988) *Current Protocols in Molecular Biology*, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

EXAMPLE 7

Growth of Genetically Modified *Corynebacterium Glutamicum*—Media and Culture Conditions

[0179] Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for *Corynebacteria* are both well-known and readily available (Lieb et al. (1989) *Appl. Microbiol. Biotechnol.*, 32: 205-210; von der Osten et al. (1998) *Biotechnology Letters*, 11: 11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: The Prokaryotes, Volume II, Balows, A. et al., eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace elements. Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts, such as NH₄Cl or (NH₄)₂SO₄, NH₄OH, nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

[0180] Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate-salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuic acid, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (eds. P. M. Rhodes, P. F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3). It is also

possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFCO) or others.

[0181] All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121° C.) or by sterile filtration. The components can either be sterilized together or, if necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

[0182] Culture conditions are defined separately for each experiment. The temperature should be in a range between 15° C. and 45° C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH₄OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the microorganisms, the pH can also be controlled using gaseous ammonia.

[0183] The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100-300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

[0184] If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD₆₀₀ of 0.5-1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2.5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30° C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

EXAMPLE 8

In Vitro Analysis of the Function of Mutant Proteins

[0185] The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments

to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one of ordinary skill in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E. C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N. C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P. D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H. U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

[0186] The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. et al. (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

[0187] The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R. B. (1989) "Pores, Channels and Transporters", in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

EXAMPLE 9

Analysis of Impact of Mutant Protein on the Production of the Desired Product

[0188] The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing the medium and/or the cellular component for increased production of the desired product (i.e., an amino acid). Such analysis techniques are well known to one of ordinary skill in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. et al., (1987) "Applications of HPLC in Biochemistry" in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al. (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P. A. et al. (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J. F. and Cabral, J. M. S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaei-

witz, J. A. and Henry, J. D. (1988) Biochemical separations, in: Ullmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F. J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

[0189] In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (e.g., sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P. M. Rhodes and P. F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

EXAMPLE 10

Purification of the Desired Product from *C. glutamicum* Culture

[0190] Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum* cells, then the cells are removed from the culture by low-speed centrifugation, and the supernatant fraction is retained for further purification.

[0191] The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One of ordinary skill in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

[0192] There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J. E. & Ollis, D. F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

[0193] The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC),

spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994) *Appl. Environ. Microbiol.* 60: 133-140; Malakhova et al. (1996) *Biotehnologiya* 11: 27-32; and Schmidt et al. (1998) *Bioprocess Engineer.* 19: 67-70. Ullmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

EXAMPLE 11

Analysis of the Gene Sequences of the Invention

[0194] The comparison of sequences and determination of percent homology between two sequences are art-known techniques, and can be accomplished using a mathematical algorithm, such as the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87: 2264-68, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-77. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215: 403-10. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to HA nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to HA protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17): 3389-3402. When utilizing BLAST and Gapped BLAST programs, one of ordinary skill in the art will know how to optimize the parameters of the program (e.g., XBLAST and NBLAST) for the specific sequence being analyzed.

[0195] Another example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Meyers and Miller ((1988) *Comput. Appl. Biosci.* 4: 11-17). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art, and include ADVANCE and ADAM, described in Torelli and Robotti (1994) *Comput. Appl. Biosci.* 10: 3-5; and FASTA, described in Pearson and Lipman (1988) *P.N.A.S.* 85: 2444-8.

[0196] The percent homology between two amino acid sequences can also be accomplished using the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. The percent homology between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package, using standard parameters, such as a gap weight of 50 and a length weight of 3.

[0197] A comparative analysis of the gene sequences of the invention with those present in Genbank has been

performed using techniques known in the art (see, e.g., Bexevanis and Ouellette, eds. (1998) *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. John Wiley and Sons: New York). The gene sequences of the invention were compared to genes present in Genbank in a three-step process. In a first step, a BLASTN analysis (e.g., a local alignment analysis) was performed for each of the sequences of the invention against the nucleotide sequences present in Genbank, and the top 500 hits were retained for further analysis. A subsequent FASTA search (e.g., a combined local and global alignment analysis, in which limited regions of the sequences are aligned) was performed on these 500 hits. Each gene sequence of the invention was subsequently globally aligned to each of the top three FASTA hits, using the GAP program in the GCG software package (using standard parameters). In order to obtain correct results, the length of the sequences extracted from Genbank were adjusted to the length of the query sequences by methods well-known in the art. The results of this analysis are set forth in Table 4. The resulting data is identical to that which would have been obtained had a GAP (global) analysis alone been performed on each of the genes of the invention in comparison with each of the references in Genbank, but required significantly reduced computational time as compared to such a database-wide GAP (global) analysis. Sequences of the invention for which no alignments above the cutoff values were obtained are indicated on Table 4 by the absence of alignment information. It will further be understood by one of ordinary skill in the art that the GAP alignment homology percentages set forth in Table 4 under the heading "% homology (GAP)" are listed in the European numerical format, wherein a ',' represents a decimal point. For example, a value of "40,345" in this column represents "40.345%".

EXAMPLE 12

Construction and Operation of DNA Microarrays

[0198] The sequences of the invention may additionally be used in the construction and application of DNA microarrays (the design, methodology, and uses of DNA arrays are well known in the art, and are described, for example, in Schena, M. et al. (1995) *Science* 270: 467-470; Wodicka, L. et al. (1997) *Nature Biotechnology* 15: 1359-1367; DeSaizieu, A. et al. (1998) *Nature Biotechnology* 16: 45-48; and DeRisi, J. L. et al. (1997) *Science* 278: 680-686).

[0199] DNA microarrays are solid or flexible supports consisting of nitrocellulose, nylon, glass, silicone, or other materials. Nucleic acid molecules may be attached to the surface in an ordered manner. After appropriate labeling, other nucleic acids or nucleic acid mixtures can be hybridized to the immobilized nucleic acid molecules, and the label may be used to monitor and measure the individual signal intensities of the hybridized molecules at defined regions. This methodology allows the simultaneous quantification of the relative or absolute amount of all or selected nucleic acids in the applied nucleic acid sample or mixture. DNA microarrays, therefore, permit an analysis of the expression of multiple (as many as 6800 or more) nucleic acids in parallel (see, e.g., Schena, M. (1996) *BioEssays* 18(5): 427-431).

[0200] The sequences of the invention may be used to design oligonucleotide primers which are able to amplify

defined regions of one or more *C. glutamicum* genes by a nucleic acid amplification reaction such as the polymerase chain reaction. The choice and design of the 5' or 3' oligonucleotide primers or of appropriate linkers allows the covalent attachment of the resulting PCR products to the surface of a support medium described above (and also described, for example, Schena, M. et al. (1995) *Science* 270: 467-470).

[0201] Nucleic acid microarrays may also be constructed by *in situ* oligonucleotide synthesis as described by Wodicka, L. et al. (1997) *Nature Biotechnology* 15: 1359-1367. By photolithographic methods, precisely defined regions of the matrix are exposed to light. Protective groups which are photolabile are thereby activated and undergo nucleotide addition, whereas regions that are masked from light do not undergo any modification. Subsequent cycles of protection and light activation permit the synthesis of different oligonucleotides at defined positions. Small, defined regions of the genes of the invention may be synthesized on microarrays by solid phase oligonucleotide synthesis.

[0202] The nucleic acid molecules of the invention present in a sample or mixture of nucleotides may be hybridized to the microarrays. These nucleic acid molecules can be labeled according to standard methods. In brief, nucleic acid molecules (e.g., mRNA molecules or DNA molecules) are labeled by the incorporation of isotopically or fluorescently labeled nucleotides, e.g., during reverse transcription or DNA synthesis. Hybridization of labeled nucleic acids to microarrays is described (e.g., in Schena, M. et al. (1995) *supra*; Wodicka, L. et al. (1997), *supra*; and DeSaizieu A. et al. (1998), *supra*). The detection and quantification of the hybridized molecule are tailored to the specific incorporated label. Radioactive labels can be detected, for example, as described in Schena, M. et al. (1995) *supra*) and fluorescent labels may be detected, for example, by the method of Shalon et al. (1996) *Genome Research* 6: 639-645).

[0203] The application of the sequences of the invention to DNA microarray technology, as described above, permits comparative analyses of different strains of *C. glutamicum* or other *Corynebacteria*. For example, studies of inter-strain variations based on individual transcript profiles and the identification of genes that are important for specific and/or desired strain properties such as pathogenicity, productivity and stress tolerance are facilitated by nucleic acid array methodologies. Also, comparisons of the profile of expression of genes of the invention during the course of a fermentation reaction are possible using nucleic acid array technology.

EXAMPLE 13

Analysis of the Dynamics of Cellular Protein Populations

[0204] (Proteomics)

[0205] The genes, compositions, and methods of the invention may be applied to study the interactions and dynamics of populations of proteins, termed 'proteomics'. Protein populations of interest include, but are not limited to, the total protein population of *C. glutamicum* (e.g., in comparison with the protein populations of other organisms), those proteins which are active under specific environmental or metabolic conditions (e.g., during fermenta-

tion, at high or low temperature, or at high or low pH), or those proteins which are active during specific phases of growth and development.

[0206] Protein populations can be analyzed by various well-known techniques, such as gel electrophoresis. Cellular proteins may be obtained, for example, by lysis or extraction, and may be separated from one another using a variety of electrophoretic techniques. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins largely on the basis of their molecular weight. Isoelectric focusing polyacrylamide gel electrophoresis (IEF-PAGE) separates proteins by their isoelectric point (which reflects not only the amino acid sequence but also posttranslational modifications of the protein). Another, more preferred method of protein analysis is the consecutive combination of both IEF-PAGE and SDS-PAGE, known as 2-D-gel electrophoresis (described, for example, in Hermann et al. (1998) *Electrophoresis* 19: 3217-3221; Fountoulakis et al. (1998) *Electrophoresis* 19: 1193-1202; Langen et al. (1997) *Electrophoresis* 18: 1184-1192; Antelmann et al. (1997) *Electrophoresis* 18: 1451-1463). Other separation techniques may also be utilized for protein separation, such as capillary gel electrophoresis; such techniques are well known in the art.

[0207] Proteins separated by these methodologies can be visualized by standard techniques, such as by staining or labeling. Suitable stains are known in the art, and include Coomassie Brilliant Blue, silver stain, or fluorescent dyes such as Sypro Ruby (Molecular Probes). The inclusion of radioactively labeled amino acids or other protein precursors (e.g., ^{35}S -methionine, ^{35}S -cysteine, ^{14}C -labelled amino acids, ^{15}N -amino acids, $^{15}\text{NO}_3^-$ or $^{15}\text{NH}_4^+$ or ^{13}C -labelled amino acids) in the medium of *C. glutamicum* permits the labeling of proteins from these cells prior to their separation. Similarly, fluorescent labels may be employed. These labeled proteins can be extracted, isolated and separated according to the previously described techniques.

[0208] Proteins visualized by these techniques can be further analyzed by measuring the amount of dye or label used. The amount of a given protein can be determined quantitatively using, for example, optical methods and can be compared to the amount of other proteins in the same gel or in other gels. Comparisons of proteins on gels can be made, for example, by optical comparison, by spectroscopy, by image scanning and analysis of gels, or through the use of photographic films and screens. Such techniques are well-known in the art.

[0209] To determine the identity of any given protein, direct sequencing or other standard techniques may be employed. For example, N- and/or C-terminal amino acid sequencing (such as Edman degradation) may be used, as may mass spectrometry (in particular MALDI or ESI techniques (see, e.g., Langen et al. (1997) *Electrophoresis* 18: 1184-1192)). The protein sequences provided herein can be used for the identification of *C. glutamicum* proteins by these techniques.

[0210] The information obtained by these methods can be used to compare patterns of protein presence, activity, or modification between different samples from various biological conditions (e.g., different organisms, time points of

fermentation, media conditions, or different biotopes, among others). Data obtained from such experiments alone, or in combination with other techniques, can be used for various applications, such as to compare the behavior of various organisms in a given (e.g., metabolic) situation, to increase the productivity of strains which produce fine chemicals or to increase the efficiency of the production of fine chemicals.

[0211] Equivalents

[0212] Those of ordinary skill in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

TABLE 1

Genes in the Application

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identifi- cation Code	Contig.	NT Start	NT Stop	Function
1 2	RXA02702	GR00758	1572	115	UDP-N-ACETYLMURAMATE-ALANINE LIGASE (EC 6.3.2.8)	
3 4	RXA02705	GR00758	5803	4388	UDP-N-ACETYLMURAMOYLALANINE-D-GLUTAMATE LIGASE (EC 6.3.2.9)	
5 6	RXA01254	GR00365	3807	2539	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMATE-2,6-DIAMINOPIMELATE LIGASE (EC 6.3.2.13)	
7 8	RXN02707	VV0017	20110	18581	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE-D-ALANYL-D-ALANYL LIGASE	
9 10	F RXA02707	GR00758	7264	6920	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE-D-ALANYL-D-ALANYL LIGASE (EC 6.3.2.15)	
11 12	F RXA02708	GR00758	7694	7260	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE-D-ALANYL-D-ALANYL LIGASE (EC 6.3.2.15)	
13 14	F RXA02709	GR00758	8451	7723	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE-D-ALANYL-D-ALANYL LIGASE (EC 6.3.2.15)	
15 16	RXA02710	GR00758	10035	8473	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMATE-2,6-DIAMINOPIMELATE LIGASE (EC 6.3.2.13)	
17 18	RXN00531	VV0079	19063	19557	FINE TANGLED PILI MAJOR SUBUNIT	
19 20	RXA00944	GR00259	1573	602	NADPH DEHYDROGENASE 3 (EC 1.6.99.1)	
21 22	RXS02560	VV0101	9922	10788	NADPH-FLAVIN OXIDOREDUCTASE (EC 1.6.99.—)	
23 24	RXS03119	VV0098	86877	87008	SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1)	
25 26	RXS03120	VV0098	87351	87476	SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1)	
<u>Cell wall biosynthesis</u>						
27 28	RXA01430	GR00417	7458	6271	N-ACETYLMURAMOYL-L-ALANINE AMIDASE (EC 3.5.1.28)	
29 30	RXA02641	GR00749	5097	3022	N-ACETYLMURAMOYL-L-ALANINE AMIDASE (EC 3.5.1.28)	
31 32	RXA00135	GR00021	1709	2962	UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.7)	
33 34	RXA02706	GR00758	6910	5813	PHOSPHO-N-ACETYLMURAMOYL-PENTAPEPTIDE-TRANSFERASE (EC 2.7.8.13)	
35 36	RXA02411	GR00703	1845	997	GLUTAMATE RACEMASE (EC 5.1.1.3)	
37 38	RXN01022	VV0143	4460	3381	D-ALANINE-D-ALANINE LIGASE (EC 6.3.2.4)	
39 40	F RXA01022	GR00292	3	806	D-ALANINE-D-ALANINE LIGASE (EC 6.3.2.4)	
41 42	RXA02703	GR00758	2698	1610	UDP-N-ACETYLGLUCOSAMINE—N-ACETYLMURAMYL-(PENTAPEPTIDE) PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE (EC 2.4.1.—)	
43 44	RXA02711	GR00758	12273	10162	PENICILLIN-BINDING PROTEIN 2	
45 46	RXA02859	GR10005	846	121	PENICILLIN-BINDING PROTEIN 5* PRECURSOR (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4)	
47 48	RXA00569	GR00152	3928	4953	PENICILLIN-BINDING PROTEIN 4	
49 50	RXN03092	VV0054	10445	9561	PENICILLIN-BINDING PROTEIN 1A	
51 52	F RXA00594	GR00158	3525	4457	PENICILLIN-BINDING PROTEIN 1A	
53 54	RXA01828	GR00516	7736	6315	PENICILLIN-BINDING PROTEIN 3	
55 56	RXA00612	GR00162	3	1187	PENICILLIN-BINDING PROTEIN 1A	
57 58	RXA01510	GR00424	15370	16650	PENICILLIN-BINDING PROTEIN 4 PRECURSOR (PBP-4) (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4)/D-ALANYL-D-ALANINE-ENDOPEPTIDASE (EC 3.4.99.—)	
59 60	RXN01608	VV0139	3536	5374	PENICILLIN-BINDING PROTEIN 5 PRECURSOR	
61 62	F RXA01608	GR00449	837	2675	(AL008883) penecillin binding protein [<i>Mycobacterium tuberculosis</i>]	
63 64	RXA01270	GR00367	21652	20498	perosamine synthetase	
65 66	RXN00549	VV0079	31746	33419	PENICILLIN-BINDING PROTEIN 1A	
67 68	RXN00550	VV0079	33457	33777	PENICILLIN-BINDING PROTEIN 1A	
69 70	RXN03091	VV0054	9515	8970	PENICILLIN-BINDING PROTEIN 1A	
71 72	RXN03178	VV0334	921	121	PENICILLIN-BINDING PROTEIN 5* PRECURSOR (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4)	

TABLE 1-continued

<u>Genes in the Application</u>							
Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function	
73	74	F RXA02859	GR10005	846	121	PENICILLIN-BINDING PROTEIN 5* PRECURSOR (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4)	
75	76	RXN01267	VV0009	17895	16582	UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.7)	
77	78	RXN00045	VV0119	4409	5317	UDP-N-acetylglucosamine-2-epimerase (EC 5.1.3.14)/N-acetylmannosamine kinase (EC 2.7.1.60)	
						<u>Cell division</u>	
79	80	RXN02704	VV0017	16043	14355	CELL DIVISION PROTEIN FTSW	
81	82	F RXA02704	GR00758	4382	2694	CELL DIVISION PROTEIN FTSW	
83	84	RXA02722	GR00759	2729	1404	CELL DIVISION PROTEIN FTSZ	
85	86	RXA00009	GR00002	1545	646	CELL DIVISION PROTEIN FTSX	
87	88	RXA00010	GR00002	2248	1562	CELL DIVISION ATP-BINDING PROTEIN FTSE	
89	90	RXA00143	GR00022	6328	4847	CELL DIVISION INHIBITOR	
91	92	RXA00277	GR00043	1588	5	CELL DIVISION CONTROL PROTEIN 15 (EC 2.7.1.—)	
93	94	RXA00857	GR00233	2	1291	CELL DIVISION PROTEIN FTSK	
95	96	RXA01435	GR00418	2	871	CELL DIVISION CONTROL PROTEIN 15 (EC 2.7.1.—)	
97	98	RXA01511	GR00424	16655	17596	CELL CYCLE PROTEIN MESJ	
99	100	RXA01513	GR00424	18368	20926	CELL DIVISION PROTEIN FISH (EC 3.4.24.—)	
101	102	RXA02098	GR00630	4161	5906	CELL DIVISION PROTEIN FTSY	
103	104	RXA02713	GR00758	14077	13067	Hypothetical Cell Division Protein mraW	
105	106	RXN02723	VV0017	11745	11080	FTSQ	
107	108	F RXA02723	GR00759	3460	2984	FTSQ	
109	110	RXA01426	GR00417	2777	3403	GLUCOSE INHIBITED DIVISION PROTEIN B	
111	112	RXA01428	GR00417	4495	5631	STAGE 0 SPORULATION PROTEIN J	
113	114	RXA01640	GR00456	4661	1344	STAGE III SPORULATION PROTEIN E	
115	116	RXA01829	GR00516	9058	7736	STAGE V SPORULATION PROTEIN E	
117	118	RXA01427	GR00417	3512	4432	SOJ PROTEIN	
119	120	RXN02973	VV0229	657	4	SOJ PROTEIN	
121	122	F RXA01603	GR00447	14043	14663	SOJ PROTEIN	
123	124	RXN00818	VV0054	28524	27685	INHIBITION OF MORPHOLOGICAL DIFFERENTIATION	
						<u>Proteolysis</u>	
125	126	RXN03028	VV0008	41156	43930	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA	
127	128	F RXA02470	GR00715	2216	3196	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA	
129	130	F RXA02471	GR00715	3159	4991	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA	
131	132	RXN03094	VV0057	1794	43	CLPB PROTEIN	
133	134	F RXA01668	GR00464	2205	3920	CLPB PROTEIN	
135	136	RXN02937	VV0098	85783	85382	N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5.1.14)	
137	138	RXN03077	VV0043	1729	2913	N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5.1.14)	
139	140	F RXA02855	GR10002	1693	2877	N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5.1.14), hippurate hydrolase	
141	142	RXN00982	VV0149	7596	6091	(L42758) proteinase [Streptomyces lividans]	
143	144	F RXA00977	GR00275	1647	2660	(L42758) proteinase [Streptomyces lividans]	
145	146	F RXA00982	GR00276	5194	4949	(L42758) proteinase [Streptomyces lividans]	
147	148	RXN01181	VV0065	1	957	AMINOPEPTIDASE A/I (EC 3.4.11.1)	
149	150	F RXA01181	GR00337	1	957	AMINOPEPTIDASE	
151	152	RXN01014	VV0209	13328	10728	AMINOPEPTIDASE N (EC 3.4.11.2)	
153	154	F RXA01014	GR00289	3	1580	AMINOPEPTIDASE N (EC 3.4.11.2)	
155	156	F RXA01018	GR00290	2289	3152	AMINOPEPTIDASE N (EC 3.4.11.2)	
157	158	RXN01046	VV0015	47863	49641	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA	
159	160	RXN01974	VV0218	3793	5577	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA	
161	162	RXN01120	VV0182	5678	4401	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPX	
163	164	F RXA01120	GR00310	2349	1072	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPX	
165	166	RXN00397	VV0025	3803	4603	XAA-PRO AMINOPEPTIDASE (EC 3.4.11.9)	
167	168	RXN01868	VV0127	9980	11905	ZINC METALLOPROTEASE (EC 3.4.24.—)	
169	170	F RXA01868	GR00534	1640	30	ZINC METALLOPROTEASE (EC 3.4.24.—)	
171	172	F RXA01869	GR00534	1954	1652	ZINC METALLOPROTEASE (EC 3.4.24.—)	
173	174	RXN00499	VV0086	8158	9438	PROLINE IMINOPEPTIDASE (EC 3.4.11.5)	
175	176	F RXA00499	GR00125	3	959	PROLINE IMINOPEPTIDASE	
177	178	RXN01277	VV0009	32155	34158	PROLYL ENDOPEPTIDASE (EC 3.4.21.26)	
179	180	F RXA01277	GR00368	1738	50	PROLYL ENDOPEPTIDASE (EC 3.4.21.26)	
181	182	RXN00675	VV0005	33258	34049	METHIONINE AMINOPEPTIDASE (EC 3.4.11.18)	
183	184	F RXA00675	GR00178	2	484	METHIONINE AMINOPEPTIDASE (EC 3.4.11.18)	
185	186	RXN00877	VV0099	2221	3885	PEPTIDYL-DIPEPTIDASE DCP (EC 3.4.15.5)	
187	188	F RXA00877	GR00242	3	1067	PEPTIDYL-DIPEPTIDASE DCP (EC 3.4.15.5)	
189	190	RXN01226	VV0064	4172	4711	PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29)	
191	192	RXN01963	VV0200	689	6	Hypothetical Secretory Serine Protease (EC 3.4.21.—)	

TABLE 1-continued

<u>Genes in the Application</u>								
Nucleic Acid	Amino Acid	SEQ ID NO	SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
193	194	RXN00621	VV0135	5853	5071	PROTEASE II (EC 3.4.21.83)		
195	196	F RXA00621	GR00163	4075	4857	PTRB periplasmic protease		
197	198	RXN00622	VV0135	5150	3735	PROTEASE II (EC 3.4.21.83)		
199	200	F RXA00622	GR00163	4778	6193	PTRB periplasmic protease		
201	202	RXN02146	VV0300	14742	15368	PROTEIN P60 PRECURSOR		
203	204	RXN03133	VV0127	39393	40076	HYDROGENASE 1 MATURATION PROTEASE (EC 3.4.—.—)		
205	206	RXN02820	VV0131	4799	6109	GAMMA-GLUTAMYLTRANSEPTIDASE (EC 2.3.2.2)		
207	208	F RXA02820	GR00801	1	507	GAMMA-GLUTAMYLTRANSEPTIDASE (EC 2.3.2.2)		
209	210	F RXA02000	GR00589	3430	3933	GAMMA-GLUTAMYLTRANSEPTIDASE (EC 2.3.2.2)		
211	212	RXN02944	VV0169	12751	12074	GAMMA-GLUTAMYLTRANSEPTIDASE PRECURSOR (EC 2.3.2.2)		
213	214	RXS00197	VV0115	2733	1522	Membrane Spanning Protease		
215	216	RXS01223	VV0064	7528	8139	PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29)		
217	218	RXS01642	VV0005	49423	48182	Serine protease		
<u>Enzymes in general</u>								
219	220	RXA01728	GR00489	2452	1478	BETA C-S LYASE (EC 3.—.—.—) PUTATIVE AMINOTRANSFERASE		
221	222	RXA02214	GR00650	954	1562	Acetyltransferases		
223	224	RXA02716	GR00758	16827	17387	Acetyltransferases		
225	226	RXN01499	VV0008	7034	3213	ENTEROBACTIN SYNTHETASE COMPONENT F		
227	228	FRXA01499	GR00424	7034	3213	Acetyltransferases (the isoleucine patch superfamily)		
229	230	RXN00787	VV0321	3736	5637	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1.4.99.1)		
231	232	F RXA00787	GR00209	598	5	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1.4.99.1)		
233	234	F RXA00791	GR00210	831	4	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1.4.99.1)		
235	236	RXA01057	GR00296	7548	6046	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1.4.99.1)		
237	238	RXA01055	GR00296	4821	4720	D-AMINO ACID DEHYDROGENASE SMALL SUBUNIT (EC 1.4.99.1)		
239	240	RXA01056	GR00296	5952	5053	D-AMINO ACID DEHYDROGENASE SMALL SUBUNIT (EC 1.4.99.1)		
241	242	RXN02021	VV0160	2008	1061	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase (EC 2.3.1.117)		
243	244	RXS00949				quinate dehydrogenase (pyrroloquinoline-quinone) (EC 1.1.99.25)		
245	246	RXS00004	VV0196	6930	6460	NITRILASE REGULATOR		
247	248	RXS00166	VV0232	3650	4309	Methyltransferase		
249	250	RXS00288	VV0079	14586	15596	QUINONE OXIDOREDUCTASE (EC 1.6.5.5)		
251	252	RXS01114	VV0182	9118	10341	3-KETOACYL-COA THIOLASE (EC 2.3.1.16)		
253	254	RXS01205	VV0268	893	363	UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.—)		
255	256	RXS01269	VV0009	21430	20990	UNDECAPRENYL-PHOSPHATE GALACTOSEPHOSPHOTRANSFERASE (EC 2.7.8.6)		
257	258	RXS01421	VV0122	16024	15638	ACYLTRANSFERASE (EC 2.3.1.—)		
259	260	RXS01491	VV0139	36800	37450	DNA FOR L-PROLINE 3-HYDROXYLASE, COMPLETE CDS		
261	262	RXS01572	VV0009	43945	44436	ALCOHOL DEHYDROGENASE (EC 1.1.1.1)		
263	264	RXS02453	VV0107	7370	8122	ACETOIN(DIACETYL) REDUCTASE (EC 1.1.1.5)		
265	266	RXS02474	VV0008	47021	46248	(S,S)-butane-2,3-diol dehydrogenase (EC 1.1.1.76)		
267	268	RXS02485	VV0007	2359	3459	UDP-N-ACETYLENOLPYRUVOYLG LUCOSAMINE REDUCTASE (EC 1.1.1.158)		
269	270	RXS02539	VV0057	17332	15815	ALDEHYDE DEHYDROGENASE (EC 1.2.1.3)		
271	272	RXS02578	VV0098	7668	6565	ACYLTRANSFERASE		
273	274	RXS02741	VV0074	5768	6733	QUINONE OXIDOREDUCTASE (EC 1.6.5.5)		
275	276	RXS03061	VV0034	108	437	ALDEHYDE DEHYDROGENASE (EC 1.2.1.3)		
277	278	RXS03150	VV0155	10678	10055	ALDEHYDE DEHYDROGENASE (EC 1.2.1.3)		
279	280	RXS02554				Oxidoreductase (EC 1.1.1.—)		
281	282	RXS03058				METHYLTRANSFERASE (EC 2.1.1.—)		
283	284	RXS03218				CAFFEOYL-COA O-METHYLTRANSFERASE (EC 2.1.1.104)		
285	286	F RXA01918	GR00549	4644	5057	CAFFEOYL-COA O-METHYLTRANSFERASE (EC 2.1.1.104)		
287	288	RXC00110	VV0054	27517	26969	Protein involved in hydrolysis of epoxides		
289	290	RXC01971	VV0105	4545	3715	Metal-Dependent Hydrolase		
<u>Genes encoding enzymes for the metabolism of inorganic compounds</u>								
<u>Phosphate and Phosphonate metabolism</u>								
291	292	RXA02118	GR00636	2124	1783	PHNA PROTEIN		
293	294	RXA00078	GR00012	6375	5962	PHNB PROTEIN		
295	296	RXA02105	GR00632	294	4	PHNB PROTEIN		
297	298	RXN00663	VV0142	10120	11493	PHOH PROTEIN HOMOLOG		
299	300	F RXA00663	GR00173	1222	227	PHOH PROTEIN HOMOLOG		
301	302	RXA00888	GR00242	14325	15341	PHOH PROTEIN HOMOLOG		
303	304	RXA01437	GR00418	3932	2550	PHOSPHATE ACETYLTRANSFERASE (EC 2.3.1.8)		
305	306	RXN00778	VV0103	18126	19250	PHOSPHATE-BINDING PERIPLASMIC PROTEIN PRECURSOR		
307	308	F RXA00778	GR00205	9079	8246	PHOSPHATE-BINDING PERIPLASMIC PROTEIN PRECURSOR		
309	310	RXA02497	GR00720	10059	10985	EXOPOLYPHOSPHATASE (EC 3.6.1.11)		
311	312	RXA01477	GR00422	8469	10016	ALKALINE PHOSPHATASE D PRECURSOR (EC 3.1.3.1)		

TABLE 1-continued

<u>Genes in the Application</u>						
Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
313 314	RXA01509	GR00424	15169	14696	INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)	
315 316	RXA00100	GR00014	9512	10111	DEDA PROTEIN, similar to alkaline phosphatase	
317 318	RXA00615	GR00162	3355	2774	DEDA PROTEIN	
319 320	RXN00250	VV0189	286	1032	DEDA PROTEIN - ALKALINE PHOSPHATASE LIKE PROTEIN	
321 322	F RXA02010	GR00602	79	525	DEDA PROTEIN	
323 324	RXA02120	GR00636	5021	4260	CARBOXYVINYL-CARBOXYPHOSPHONATE PHOSPHORYLMUTASE (EC 2.7.8.23)	
325 326	RXS01000	VV0106	7252	6407	PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE	
327 328	RXS01002	VV0106	8858	8055	PHOSPHONATES TRANSPORT ATP-BINDING PROTEIN PHNC	
329 330	RXS01003	VV0106	8055	7252	PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE	
331 332	RXS01902	VV0098	84095	83037	alkaline phosphatase	
<u>Fe-Metabolism</u>						
333 334	RXA01967	GR00567	1848	706	FERRIC ENTEROCHELIN ESTERASE HOMOLOG	
335 336	RXA00070	GR00011	3436	3867	FERRIC UPTAKE REGULATION PROTEIN	
337 338	RXA01934	GR00555	7192	7749	FERRIPYOCHELIN BINDING PROTEIN	
339 340	RXN01997	VV0084	33308	33793	FERRITIN	
341 342	F RXA01997	GR00586	546	935	FERRITIN	
343 344	RXA01082	GR00302	1486	827	IRON REPRESSOR	
345 346	RXA01236	GR00358	2185	1241	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR	
347 348	RXA01354	GR00393	2692	1757	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR	
349 350	RXA01620	GR00451	2585	3532	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR	
351 352	RXA02052	GR00624	4586	3795	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR	
353 354	RXA00372	GR00078	1653	2729	PERIPLASMIC-IRON-BINDING PROTEIN SHIB	
355 356	RXA00088	GR00013	4389	5402	FERRIC ANGUILIBACTIN-BINDING PROTEIN PRECURSOR	
357 358	RXS00156	VV0167	1342	2451	FERROCHELATASE (EC 4.99.1.1)	
359 360	RXS00624	VV0135	2018	1332	FERROCHELATASE (EC 4.99.1.1)	
<u>Modification and degradation of aromatic compounds</u>						
361 362	RXA00024	GR00003	938	1882	ARYL-ALCOHOL DEHYDROGENASE (NADP+) (EC 1.1.1.91)	
363 364	RXA02526	GR00725	4109	5314	3-CARBOXY-CIS,CIS-MUCONATE CYCLOISOMERASE (EC 5.5.1.2)	
365 366	RXN02813	VV0128	13120	14118	3-CARBOXY-CIS,CIS-MUCONATE CYCLOISOMERASE HOMOLOG (EC 5.5.1.2)	
367 368	F RXA02813	GR00794	651	10	3-CARBOXY-CIS,CIS-MUCONATE CYCLOISOMERASE HOMOLOG (EC 5.5.1.2)	
369 370	RXA01113	GR00307	1098	862	4-CARBOXYMUCONOLACTONE DECARBOXYLASE (EC 4.1.1.44)	
371 372	RXA02126	GR00637	1556	1876	4-CARBOXYMUCONOLACTONE DECARBOXYLASE (EC 4.1.1.44)	
373 374	RXA01465	GR00421	4121	2961	MUCONATE CYCLOISOMERASE (EC 5.5.1.1)	
375 376	RXA02316	GR00665	9038	8025	MUCONATE CYCLOISOMERASE (EC 5.5.1.1)	
377 378	RXA01464	GR00421	2945	2655	MUCONOLACTONE ISOMERASE (EC 5.3.3.4)	
379 380	RXA02603	GR00742	7742	8737	4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-)	
381 382	RXN02839	VV0362	3	449	4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-)	
383 384	F RXA02839	GR00832	3	419	4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-)	
385 386	RXA01502	GR00424	8385	9617	BENZENE 1,2-DIOXYGENASE SYSTEM FERREDOXIN-NAD(+) REDUCTASE COMPONENT (EC 1.18.1.3)	
387 388	RXA02828	GR00813	15	572	BIPHENYL-2,3-DIOL 1,2-DIOXYGENASE III (EC 1.13.11.39)	
389 390	RXA02064	GR00626	5223	4585	CAFFEOYL-COA O-METHYLTRANSFERASE (EC 2.1.1.104)	
391 392	RXN00639	VV0128	7858	8712	CATECHOL 1,2-DIOXYGENASE (EC 1.13.11.1)	
393 394	F RXA00639	GR00168	665	6	CATECHOL 1,2-DIOXYGENASE (EC 1.13.11.1)	
395 396	RXN01653	VV0321	12867	11407	DIBENZOTHIOPHENE DESULFURIZATION ENZYME A	
397 398	F RXA00797	GR00212	445	804	DIBENZOTHIOPHENE DESULFURIZATION ENZYME A	
399 400	F RXA01653	GR00458	1909	971	DIBENZOTHIOPHENE DESULFURIZATION ENZYME A	
401 402	RXN02530	VV0057	5469	6125	DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 1 (EC 1.14.13.8)	
403 404	F RXA02530	GR00726	20	469	DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 1 (EC 1.14.13.8)	
405 406	RXA02083	GR00629	1720	311	DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 2 (EC 1.14.13.8)	
407 408	RXA00892	GR00243	2188	1295	PARANITROBENZYL ESTERASE (EC 3.1.1.-)	
409 410	RXA02092	GR00629	12153	10516	PARANITROBENZYL ESTERASE (EC 3.1.1.-)	
411 412	RXN00658	VV0083	15705	16397	PHENOL 2-MONOXYGENASE (EC 1.14.13.7)	
413 414	F RXA00658	GR00170	321	4	PHENOL 2-MONOXYGENASE (EC 1.14.13.7)	
415 416	RXA01385	GR00406	5320	3440	PHENOL 2-MONOXYGENASE (EC 1.14.13.7)	
417 418	RXN01461	VV0128	12414	13025	PROTocatechuate 3,4-DIOXYGENASE ALPHA CHAIN (EC 1.13.11.3)	
419 420	F RXA01461	GR00421	463	5	PROTocatechuate 3,4-DIOXYGENASE ALPHA CHAIN (EC 1.13.11.3)	
421 422	RXA01462	GR00421	1167	478	PROTocatechuate 3,4-DIOXYGENASE BETA CHAIN (EC 1.13.11.3)	
423 424	RXN00641	VV0128	7440	5950	TOLUATE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12.-)	
425 426	F RXA00640	GR00168	1083	1331	TOLUATE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12.-)	

TABLE 1-continued

Genes in the Application

Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
427	428	F RXA00641	GR00168	1533	2573	TOLUATE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12.-)
429	430	RXA00642	GR00168	2616	3107	TOLUATE 1,2-DIOXYGENASE BETA SUBUNIT (EC 1.14.12.-)
431	432	RXA00643	GR00168	3122	4657	TOLUATE 1,2-DIOXYGENASE ELECTRON TRANSFER COMPONENT
433	434	RXN01993	VV0182	16	1143	VANILLATE DEMETHYLASE (EC 1.14.-.-)
435	436	F RXA01993	GR00584	1	366	VANILLATE DEMETHYLASE (EC 1.14.-.-)
437	438	F RXA02012	GR00604	2	670	VANILLATE DEMETHYLASE (EC 1.14.-.-)
439	440	RXA01994	GR00584	373	1347	VANILLATE DEMETHYLASE OXIDOREDUCTASE (EC 1.-.-.-)
441	442	RXA02535	GR00726	6599	7753	XYLENE MONOOXYGENASE ELECTRON TRANSFER COMPONENT
443	444	RXA00964	GR00269	1575	451	1-hydroxy-2-naphthoate 1,2-dioxygenase (EC 1.13.11.38)
445	446	RXN01466	VV0019	7050	6091	ARYLESTERASE (EC 3.1.1.2)
447	448	F RXA01466	GR00422	826	5	ARYLESTERASE (EC 3.1.1.2)
449	450	RXN03036	VV0014	671	6	PROTOCATECHUATE 3,4-DIOXYGENASE BETA CHAIN (EC 1.13.11.3)
451	452	F RXA02895	GR10037	671	6	CHLOROCATECHOL 1,2-DIOXYGENASE (EC 1.13.11.1)
453	454	RXA02449	GR00710	1458	2360	hydroxyquinol 1,2-dioxygenase (EC 1.13.11.37)
455	456	RXN00178	VV0174	14670	15554	hydroxyquinol 1,2-dioxygenase (EC 1.13.11.37)
457	458	F RXA00178	GR00028	304	1188	HYDROXYQUINOL-1,2-DIOXYGENASE
459	460	RXA02111	GR00632	4310	5593	QUINOLINATE SYNTHETASE A
461	462	RXA00644	GR00168	4657		CIS-1,2-DIHYDROXYCYCLOHEXA-3,5-DIENE-1-CARBOXYLATE DEHYDROGENASE (EC 1.3.1.55)
463	464	RXN00177	VV0174	13589	14656	MALEYLACETATE REDUCTASE (EC 1.3.1.32)
465	466	F RXA00177	GR00028	3	290	MALEYLACETATE REDUCTASE (EC 1.3.1.32) metabolism of 2,4,5-trichlorophenoxyacetic acid
467	468	RXA02448	GR00710	340	1428	MALEYLACETATE REDUCTASE (EC 1.3.1.32)
469	470	RXA00048	GR00008	2185	527	3-(3-HYDROXYPHENYL) PROPIONATE HYDROXYLASE
471	472	RXA01126	GR00313	2	565	POSSIBLE 2-HYDROXYHEPTA-2,4-DIENE-1,7-DIOATE ISOMERASE
473	474	RXA01117	GR00309	1713	973	SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE PRECURSOR (EC 2.8.3.5)
475	476	RXA00772	GR00205	2715	1210	SUCCINYL-COA:COENZYME A TRANSFERASE (EC 2.8.3.-)
477	478	RXA01288	GR00372	2018	1644	SUCCINYL-COA:COENZYME A TRANSFERASE (EC 2.8.3.-)

[0213]

TABLE 2

GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
A09073	ppg	Phosphoenol pyruvate carboxylase	Bachmann, B. et al. "DNA fragment coding for phosphoenolpyruvat carboxylase, recombinant DNA carrying said fragment, strains carrying the recombinant DNA and method for producing L-aminino acids using said strains," Patent: EP 0358940-A 3 Mar. 21, 1990
A45579, A45581, A45583, A45585 A45587		Threonine dehydratase	Moekkel, B. et al. "Production of L-isoleucine by means of recombinant micro-organisms with deregulated threonine dehydratase," Patent: WO 9519442-A 5 Jul. 20, 1995
AB003132	murC; ftsQ; ftsZ		Kobayashi, M. et al. "Cloning, sequencing, and characterization of the ftsZ gene from <i>coryneform bacteria</i> ," Biochem. Biophys. Res. Commun., 236(2): 383-388 (1997)
AB015023	murC; ftsQ		Wachi, M. et al. "A murC gene from <i>coryneform bacteria</i> ," Appl. Microbiol. Biotechnol., 51(2): 223-228 (1999)
AB018530	dtsR		Kimura, E. et al. "Molecular cloning of a novel gene, dtsR, which rescues the detergent sensitivity of a mutant derived from <i>Brevibacterium laevofermentum</i> ," Biosci. Biotechnol. Biochem., 60(10): 1565-1570 (1996)
AB018531	dtsR1; dtsR2		
AB020624	murl	D-glutamate racemase	
AB023377	tkt	transketolase	
AB024708	gltB; gltD	Glutamine 2-oxoglutarate aminotransferase large and small subunits	

TABLE 2-continued

<u>GENES IDENTIFIED FROM GENBANK</u>			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
AB025424	acn	aconitase	
AB027714	rep	Replication protein	
AB027715	rep; aad	Replication protein; aminoglycoside adenyltransferase	
AF005242	argC	N-acetylglutamate-5-semialdehyde dehydrogenase	
AF005635	glnA	Glutamine synthetase	
AF030405	hisF	cyclase	
AF030520	argG	Argininosuccinate synthetase	
AF031518	argF	Ornithine carbamoyltransferase	
AF036932	aroD	3-dehydroquinate dehydratase	
AF038548	pyc	Pyruvate carboxylase	
AF038651	dciAE; apt; rel	Dipeptide-binding protein; adenine phosphoribosyltransferase; GTP pyrophosphokinase	Wehmeier, L. et al. "The role of the <i>Corynebacterium glutamicum</i> rel gene in (p)ppGpp metabolism," <i>Microbiology</i> , 144: 1853-1862 (1998)
AF041436	argR	Arginine repressor	
AF045998	impA	Inositol monophosphate phosphatase	
AF048764	argH	Argininosuccinate lyase	
AF049897	argC; argJ; argB; argD; argF; argR; argG; argH	N-acetylglutamylphosphate reductase; ornithine acetyltransferase; N-acetylglutamate kinase; acetylornithine transaminase; ornithine carbamoyltransferase; arginine repressor; argininosuccinate synthase; argininosuccinate lyase	
AF050109	inhA	Enoyl-acyl carrier protein reductase	
AF050166	hisG	ATP phosphoribosyltransferase	
AF051846	hisA	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazolecarboxamide isomerase	
AF052652	metA	Homoserine O-acetyltransferase	Park, S. et al. "Isolation and analysis of metA, a methionine biosynthetic gene encoding homoserine acetyltransferase in <i>Corynebacterium glutamicum</i> ," <i>Mol. Cells.</i> , 8(3): 286-294 (1998)
AF053071	aroB	Dehydroquinase synthetase	
AF060558	hisH	Glutamine amidotransferase	
AF086704	hisE	Phosphoribosyl-ATP-pyrophosphohydrolase	
AF114233	aroA	5-enolpyruvylshikimate 3-phosphate synthase	
AF116184	panD	L-aspartate-alpha-decarboxylase precursor	Dusch, N. et al. "Expression of the <i>Corynebacterium glutamicum</i> panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in <i>Escherichia coli</i> ," <i>Appl. Environ. Microbiol.</i> , 65(4):1530-1539 (1999)
AF124518	aroD; aroE	3-dehydroquinase; shikimate dehydrogenase	
AF124600	aroC; aroK; aroB; pepQ	Chorismate synthase; shikimate kinase; 3-dehydroquinase synthase; putative cytoplasmic peptidase	
AF145897	inhA		
AF145898	inhA		
AJ001436	ectP	Transport of ectoine, glycine betaine, proline	Peter, H. et al. " <i>Corynebacterium glutamicum</i> is equipped with four secondary carriers for compatible solutes: Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," <i>J. Bacteriol.</i> , 180(22): 6005-6012 (1998)
AJ004934	dapD	Tetrahydrodipicolinate succinylase (incomplete ⁱ)	Wehrmann, A. et al. "Different modes of diaminopimelate synthesis and their role in cell wall integrity: A study with <i>Corynebacterium glutamicum</i> ," <i>J. Bacteriol.</i> , 180(12): 3159-3165 (1998)
AJ007732	ppc; secG; amt; ocd; soxA	Phosphoenolpyruvate-carboxylase; ?; high affinity ammonium uptake protein; putative ornithine-cyclodecarboxylase; sarcosine oxidase	
AJ010319	ftsY, glnB, glnD; sfp; amtP	Involved in cell division; PII protein; uridylyltransferase (uridylyl-removing enzyme); signal recognition particle; low affinity ammonium uptake protein	Jakoby, M. et al. "Nitrogen regulation in <i>Corynebacterium glutamicum</i> ; Isolation of genes involved in biochemical characterization of corresponding proteins," <i>FEMS Microbiol.</i> , 173(2): 303-310 (1999)

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AJ132968	cat	Chloramphenicol acetyl transferase	
AJ224946	mqa	L-malate: quinone oxidoreductase	Molenaar, D. et al. "Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from <i>Corynebacterium glutamicum</i> ," Eur. J. Biochem., 254(2): 395–403 (1998)
AJ238250	ndh	NADH dehydrogenase	Lichtinger, T. et al. "Biochemical and biophysical characterization of the cell wall porin of <i>Corynebacterium glutamicum</i> : The channel is formed by a low molecular mass polypeptide," Biochemistry, 37(43): 15024–15032 (1998)
AJ238703	porA	Porin	
D17429		Transposable element IS31831	Vertes, A. A. et al. "Isolation and characterization of IS31831, a transposable element from <i>Corynebacterium glutamicum</i> ," Mol. Microbiol., 11(4): 739–746 (1994)
D84102	odhA	2-oxoglutarate dehydrogenase	Usuda, Y. et al. "Molecular cloning of the <i>Corynebacterium glutamicum</i> (<i>Brevibacterium lactofermentum</i> AJ12036) odhA gene encoding a novel type of 2-oxoglutarate dehydrogenase," Microbiology, 142: 3347–3354 (1996)
E01358	hdh; hk	Homoserine dehydrogenase; homoserine kinase	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 1 Oct. 12, 1987
E01359		Upstream of the start codon of homoserine kinase gene	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 2 Oct. 12, 1987
E01375		Tryptophan operon	
E01376	tppL; tppE	Leader peptide; anthranilate synthase	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 Oct. 24, 1987
E01377		Promoter and operator regions of tryptophan operon	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 Oct. 24, 1987
E03937		Biotin-synthase	Hatakeyama, K. et al. "DNA fragment containing gene capable of coding biotin synthetase and its utilization," Patent: JP 1992278088-A 1 Oct. 02, 1992
E04040		Diamino pelargonic acid aminotransferase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 Nov. 18, 1992
E04041		Desthiobiotinsynthetase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 Nov. 18, 1992
E04307		Flavum aspartase	Kurusu, Y. et al. "Gene DNA coding aspartase and utilization thereof," Patent: JP 1993030977-A 1 Feb. 09, 1993
E04376		Isocitric acid lyase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 Mar. 09, 1993
E04377		Isocitric acid lyase N-terminal fragment	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 Mar. 09, 1993
E04484		Prephenate dehydratase	Sotouchi, N. et al. "Production of L-phenylalanine by fermentation," Patent: JP 199307635 2-A 2 Mar. 30, 1993
E05108		Aspartokinase	Fugono, N. et al. "Gene DNA coding Aspartokinase and its use," Patent: JP 1993184366-A 1 Jul. 27, 1993
E05112		Dihydro-dipicorinate synthetase	Hatakeyama, K. et al. "Gene DNA coding dihydriodicolic acid synthetase and its use," Patent: JP 1993184371-A 1 Jul. 27, 1993
E05776		Diaminopimelic acid dehydrogenase	Kobayashi, M. et al. "Gene DNA coding Diaminopimelic acid dehydrogenase and its use," Patent: JP 1993284970-A 1 Nov. 02, 1993
E05779		Threonine synthase	Kohama, K. et al. "Gene DNA coding threonine synthase and its use," Patent: JP 1993284972-A 1 Nov. 02, 1993
E06110		Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 Dec. 27, 1993
E06111		Mutated Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 Dec. 27, 1993
E06146		Acetohydroxy acid synthetase	Inui, M. et al. "Gene capable of coding Acetohydroxy acid synthetase and its use," Patent: JP 1993344893-A 1 Dec. 27, 1993
E06825		Aspartokinase	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 Mar. 08, 1994
E06826		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 Mar. 08, 1994
E06827		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 Mar. 08, 1994
E07701	secY		Honno, N. et al. "Gene DNA participating in integration of membranous protein to membrane," Patent: JP 1994169780-A 1 Jun. 21, 1994
E08177		Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 Sep. 20, 1994
E08178, E08179, E08180,		Feedback inhibition-released Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 Sep. 20, 1994

TABLE 2-continued

<u>GENES IDENTIFIED FROM GENBANK</u>			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
E08181, E08182			Inui, M. et al. "Gene DNA coding acetohydroxy acid isomeroeductase," Patent: JP 1994277067-A 1 Oct. 04, 1994
E08232		Acetohydroxy-acid isomeroeductase	
E08234	secE		Asai, Y. et al. "Gene DNA coding for translocation machinery of protein," Patent: JP 1994277073-A 1 Oct. 04, 1994
E08643		FT aminotransferase and desthiobiotin synthetase promoter region	Hatakeyama, K. et al. "DNA fragment having promoter function in <i>coryneform bacterium</i> ," Patent: JP 1995031476-A 1 Feb. 03, 1995
E08646		Biotin synthetase	Hatakeyama, K. et al. "DNA fragment having promoter function in <i>coryneform bacterium</i> ," Patent: JP 1995031476-A 1 Feb. 03, 1995
E08649		Aspartase	Kohama, K. et al. "DNA fragment having promoter function in <i>coryneform bacterium</i> ," Patent: JP 1995031478-A 1 Feb. 03, 1995
E08900		Dihydridopicolinate reductase	Madori, M. et al. "DNA fragment containing gene coding Dihydridopicolinate acid reductase and utilization thereof," Patent: JP 1995075578-A 1 Mar. 20, 1995
E08901		Diaminopimelic acid decarboxylase	Madori, M. et al. "DNA fragment containing gene coding Diaminopimelic acid decarboxylase and utilization thereof," Patent: JP 1995075579-A 1 Mar. 20, 1995
E12594		Serine hydroxymethyltransferase	Hatakeyama, K. et al. "Production of L-tryptophan," Patent: JP 1997028391-A 1 Feb. 04, 1997
E12760, E12759, E12758		transposase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
E12764		Arginyl-tRNA synthetase; diaminopimelic acid decarboxylase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
E12767		Dihydridopicolinic acid synthetase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
E12770		aspartokinase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
E12773		Dihydridopicolinic acid reductase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
E13655		Glucose-6-phosphate dehydrogenase	Hatakeyama, K. et al. "Glucose-6-phosphate dehydrogenase and DNA capable of coding the same," Patent: JP 1997224661-A 1 Sep. 02, 1997
L01508	IlvA	Threonine dehydratase	Moekel, B. et al. "Functional and structural analysis of the threonine dehydratase of <i>Corynebacterium glutamicum</i> ," J. Bacteriol., 174: 8065-8072 (1992)
L07603	EC 4.2.1.15	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	Chen, C. et al. "The cloning and nucleotide sequence of <i>Corynebacterium glutamicum</i> 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene," FEMS Microbiol. Lett., 107: 223-230 (1993)
L09232	IlvB; ilvN; ilvC	Acetohydroxy acid synthase large subunit; Acetohydroxy acid synthase small subunit; Acetohydroxy acid isomeroeductase	Keilhauer, C. et al. "Isoleucine synthesis in <i>Corynebacterium glutamicum</i> : molecular analysis of the ilvB-ilvN-ilvC operon," J. Bacteriol., 175(17): 5595-5603 (1993)
L18874	PtsM	Phosphoenolpyruvate sugar phosphotransferase	Fouet, A et al. "Bacillus subtilis sucrose-specific enzyme II of the phosphotransferase system: expression in <i>Escherichia coli</i> and homology to enzymes II from enteric bacteria," PNAS USA, 84(24): 8773-8777 (1987); Lee, J. K. et al. "Nucleotide sequence of the gene encoding the <i>Corynebacterium glutamicum</i> mannose enzyme II and analyses of the deduced protein sequence," FEMS Microbiol. Lett., 119(1-2): 137-145 (1994)
L27123	aceB	Malate synthase	Lee, H-S. et al. "Molecular characterization of aceB, a gene encoding malate synthase in <i>Corynebacterium glutamicum</i> ," J. Microbiol. Biotechnol., 4(4): 256-263 (1994)
L27126		Pyruvate kinase	Jetten, M. S. et al. "Structural and functional analysis of pyruvate kinase from <i>Corynebacterium glutamicum</i> ," Appl. Environ. Microbiol., 60(7): 2501-2507 (1994)
L28760 L35906	aceA dtxR	Isocitrate lyase Diphtheria toxin repressor	Oguiza, J. A. et al. "Molecular cloning, DNA sequence analysis, and characterization of the <i>Corynebacterium diphtheriae</i> dtxR from <i>Brevibacterium lactofermentum</i> ," J. Bacteriol., 177(2): 465-467 (1995)
M13774		Prephenate dehydratase	Follettie, M. T. et al. "Molecular cloning and nucleotide sequence of the <i>Corynebacterium glutamicum</i> pheA gene," J. Bacteriol., 167: 695-702 (1986)
M16175	5S rRNA		Park, Y-H. et al. "Phylogenetic analysis of the <i>coryneform</i> bacteria by 56 rRNA sequences," J. Bacteriol., 169: 1801-1806 (1987)
M16663	trpE	Anthranilate synthase, 5' end	Sano, K. et al. "Structure and function of the trp operon control regions of <i>Brevibacterium lactofermentum</i> , a glutamic-acid-producing bacterium," Gene, 52: 191-200 (1987)
M16664	trpA	Tryptophan synthase, 3'end	Sano, K. et al. "Structure and function of the trp operon control regions of <i>Brevibacterium lactofermentum</i> , a glutamic-acid-producing bacterium," Gene, 52: 191-200 (1987)

TABLE 2-continued

<u>GENES IDENTIFIED FROM GENBANK</u>			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
M25819		Phosphoenolpyruvate carboxylase	O'Regan, M. et al. "Cloning and nucleotide sequence of the Phosphoenolpyruvate carboxylase-coding gene of <i>Corynebacterium glutamicum</i> ATCC13032," <i>Gene</i> , 77(2): 237-251 (1989)
M85106		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G + C content are characterized by a common insertion within their 23S rRNA genes," <i>J. Gen. Microbiol.</i> , 138: 1167-1175 (1992)
M85107, M85108		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G + C content are characterized by a common insertion within their 23S rRNA genes," <i>J. Gen. Microbiol.</i> , 138: 1167-1175 (1992)
M89931	aecD; brnQ; yhw	Beta C-S lyase; branched-chain amino acid uptake carrier; hypothetical protein yhw	Rossol, I. et al. "The <i>Corynebacterium glutamicum</i> aecD gene encodes a C-S lyase with alpha, beta-elimination activity that degrades aminoethylcysteine," <i>J. Bacteriol.</i> , 174(9): 2968-2977 (1992); Tauch, A. et al. "Isoleucine uptake in <i>Corynebacterium glutamicum</i> ATCC 13032 is directed by the brnQ gene product," <i>Arch. Microbiol.</i> , 169(4): 303-312 (1998)
S59299	trp	Leader gene (promoter)	Herry, D. M. et al. "Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of <i>Corynebacterium glutamicum</i> : identification of a mutation in the trp leader sequence," <i>Appl. Environ. Microbiol.</i> , 59(3): 791-799 (1993)
U11545	trpD	Anthraniilate phosphoribosyltransferase	O'Gara, J. P. and Duncan, L. K. (1994) Complete nucleotide sequence of the <i>Corynebacterium glutamicum</i> ATCC 21850 tpD gene." Thesis, Microbiology Department, University College Galway, Ireland.
U13922	cglIM; cglIR; cglIIR	Putative type II 5-cytosine methyltransferase; putative type II restriction endonuclease; putative type I or type III restriction endonuclease	Schafer, A. et al. "Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from <i>Corynebacterium glutamicum</i> ATCC 13032 and analysis of its role in intergeneric conjugation with <i>Escherichia coli</i> ," <i>J. Bacteriol.</i> , 176(23): 7309-7319 (1994); Schafer, A. et al. "The <i>Corynebacterium glutamicum</i> cglIM gene encoding a 5-cytosine in an MrBC-deficient <i>Escherichia coli</i> strain," <i>Gene</i> , 203(2): 95-101 (1997)
U14965 U31224	recA ppx		Ankri, S. et al. "Mutations in the <i>Corynebacterium glutamicum</i> proline biosynthetic pathway: A natural bypass of the proA step," <i>J. Bacteriol.</i> , 178(15): 4412-4419 (1996)
U31225	proC	L-proline: NADP+ 5-oxidoreductase	Ankri, S. et al. "Mutations in the <i>Corynebacterium glutamicum</i> proline biosynthetic pathway: A natural bypass of the proA step," <i>J. Bacteriol.</i> , 178(15): 4412-4419 (1996)
U31230	obg; proB; unkdh	?; gamma glutamyl kinase; similar to D-isomer specific 2-hydroxyacid dehydrogenases	Ankri, S. et al. "Mutations in the <i>Corynebacterium glutamicum</i> proline biosynthetic pathway: A natural bypass of the proA step," <i>J. Bacteriol.</i> , 178(15): 4412-4419 (1996)
U31281	bioB	Biotin synthase	Serebriskii, I. G., "Two new members of the bio B superfamily: Cloning, sequencing and expression of bio B genes of <i>Methylobacillus flagellatum</i> and <i>Corynebacterium glutamicum</i> ," <i>Gene</i> , 175: 15-22 (1996)
U35023	thtR; accBC	Thiosulfate sulfurtransferase; acyl CoA carboxylase	Jager, W. et al. "A <i>Corynebacterium glutamicum</i> gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins," <i>Arch. Microbiol.</i> , 166(2): 76-82 (1996)
U43535	cmr	Multidrug resistance protein	Jager, W. et al. "A <i>Corynebacterium glutamicum</i> gene conferring multidrug resistance in the heterologous host <i>Escherichia coli</i> ," <i>J. Bacteriol.</i> , 179(7): 2449-2451 (1997)
U43536 U53587 U89648	cipB aphA-3	Heat shock ATP-binding protein 3'5"-aminoglycoside phosphotransferase <i>Corynebacterium glutamicum</i> unidentified sequence involved in histidine biosynthesis, partial sequence	
X04960	trpA; trpB; trpC; trpD; trpE; trpG; trpL	Tryptophan operon	Matsui, K. et al. "Complete nucleotide and deduced amino acid sequences of the <i>Brevibacterium lactofermentum</i> tryptophan operon," <i>Nucleic Acids Res.</i> , 14(24): 10113-10114 (1986)
X07563	lys A	DAP decarboxylase (meso-diaminopimelate decarboxylase, EC 4.1.1.20)	Yeh, P. et al. "Nucleic sequence of the lysA gene of <i>Corynebacterium glutamicum</i> and possible mechanisms for modulation of its expression," <i>Mol. Gen. Genet.</i> , 212(1): 112-119 (1988)
X14234	EC 4.1.1.31	Phosphoenolpyruvate carboxylase	Eikmanns, B. J. et al. "The Phosphoenolpyruvate carboxylase gene of <i>Corynebacterium glutamicum</i> : Molecular cloning, nucleotide sequence, and expression," <i>Mol. Gen. Genet.</i> , 218(2): 330-339 (1989); Lepiniec, L. et al. "Sorghum Phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution," <i>Plant. Mol. Biol.</i> , 21 (3): 487-502 (1993)
X17313	fda	Fructose-biphosphate aldolase	Von der Osten, C. H. et al. "Molecular cloning, nucleotide sequence and fine-structural analysis of the <i>Corynebacterium glutamicum</i> fda gene: structural comparison of <i>C. glutamicum</i> fructose-1,6-biphosphate aldolase to class I and class II aldolases," <i>Mol. Microbiol.</i> ,
X53993	dapA	L-2,3-dihydridopicolinate synthetase (EC 4.2.1.52)	Bonnassie, S. et al. "Nucleic sequence of the dapA gene from <i>Corynebacterium glutamicum</i> ," <i>Nucleic Acids Res.</i> , 18(21): 6421 (1990)

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X54223		AttB-related site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of <i>Corynebacterium diphtheriae</i> , <i>Corynebacterium ulcerans</i> , <i>Corynebacterium glutamicum</i> , and the attP site of lambdacorynephage," FEMS. Microbiol. Lett., 66: 299-302 (1990)
X54740	argS; lysA	Arginyl-tRNA synthetase; Diaminopimelate decarboxylase	Marcel, T. et al. "Nucleotide sequence and organization of the upstream region of the <i>Corynebacterium glutamicum</i> lysA gene," Mol. Microbiol., 4(11): 1819-1830 (1990)
X55994	trpL; trpE	Putative leader peptide; anthranilate synthase component 1	Heery, D. M. et al. "Nucleotide sequence of the <i>Corynebacterium glutamicum</i> trpE gene," Nucleic Acids Res., 18(23): 7138 (1990)
X56037	thrC	Threonine synthase	Han, K. S. et al. "The molecular structure of the <i>Corynebacterium glutamicum</i> threonine synthase gene," Mol. Microbiol., 4(10): 1693-1702 (1990)
X56075	attB-related site	Attachment site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of <i>Corynebacterium diphtheriae</i> , <i>Corynebacterium ulcerans</i> , <i>Corynebacterium glutamicum</i> , and the attP site of lambdacorynephage," FEMS. Microbiol. Lett., 66: 299-302 (1990)
X57226	lysC-alpha; lysC-beta; asd	Aspartokinase-alpha subunit; Aspartokinase-beta subunit; aspartate beta semialdehyde dehydrogenase	Kalinowski, J. et al. "Genetic and biochemical analysis of the Aspartokinase from <i>Corynebacterium glutamicum</i> ," Mol. Microbiol., 5(5): 1197-1204 (1991); Kalinowski, J. et al. "Aspartokinase genes lysC alpha and lysC beta overlap and are adjacent to the aspartate beta-semialdehyde dehydrogenase gene asd in <i>Corynebacterium glutamicum</i> ," Mol. Gen. Genet., 224(3): 317-324 (1990)
X59403	gap; pgk; tpi	Glyceraldehyde-3-phosphate; phosphoglycerate kinase; triosephosphate isomerase	Eikmanns, B. J. "Identification, sequence analysis, and expression of a <i>Corynebacterium glutamicum</i> gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomeraser," J. Bacteriol., 174(19): 6076-6086 (1992)
X59404	gdh	Glutamate dehydrogenase	Bormann, E. R. et al. "Molecular analysis of the <i>Corynebacterium glutamicum</i> gdh gene encoding glutamate dehydrogenase," Mol. Microbiol., 6(3): 317-326 (1992)
X60312	lysI	L-lysine permease	Seep-Feldhaus, A. H. et al. "Molecular analysis of the <i>Corynebacterium glutamicum</i> lysI gene involved in lysine uptake," Mol. Microbiol., 5(12): 2995-3005 (1991)
X66078	csp1	Ps1 protein	Joliff, G. et al. "Cloning and nucleotide sequence of the csp1 gene encoding PS1, one of the two major secreted proteins of <i>Corynebacterium glutamicum</i> : The deduced N-terminal region of PS1 is similar to the Mycobacterium antigen 85 complex," Mol. Microbiol., 6(16): 2349-2362 (1992)
X66112	glt	Citrate synthase	Eikmanns, B. J. et al. "Cloning sequence, expression and transcriptional analysis of the <i>Corynebacterium glutamicum</i> gltA gene encoding citrate synthase," Microbiol., 140: 1817-1828 (1994)
X67737 X69103	dapB csp2	Dihydrodipicolinate reductase Surface layer protein PS2	Peyret, J. L. et al. "Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in <i>Corynebacterium glutamicum</i> ," Mol. Microbiol., 9(1): 97-109 (1993)
X69104		IS3 related insertion element	Bonamy, C. et al. "Identification of IS1206, a <i>Corynebacterium glutamicum</i> IS3-related insertion sequence and phylogenetic analysis," Mol. Microbiol., 14(3): 571-581 (1994)
X70959	leuA	Isopropylmalate synthase	Patek, M. et al. "Leucine synthesis in <i>Corynebacterium glutamicum</i> : enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis," Appl. Environ. Microbiol., 60(1): 133-140 (1994)
X71489	icd	Isocitrate dehydrogenase (NADP+)	Eikmanns, B. J. et al. "Cloning sequence analysis, expression, and inactivation of the <i>Corynebacterium glutamicum</i> icd gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme," J. Bacteriol., 177(3): 774-782 (1995)
X72855 X75083, X70584	GDHA mtrA	Glutamate dehydrogenase (NADP+) 5-methyltryptophan resistance	Heery, D. M. et al. "A sequence from a tryptophan-hyperproducing strain of <i>Corynebacterium glutamicum</i> encoding resistance to 5-methyltryptophan," Biochem. Biophys. Res. Commun., 201(3): 1255-1262 (1994)
X75085	recA		Fitzpatrick, R. et al. "Construction and characterization of recA mutant strains of <i>Corynebacterium glutamicum</i> and <i>Brevibacterium lactofermentum</i> ," Appl. Microbiol. Biotechnol., 42(4): 575-580 (1994)
X75504	aceA; thiX	Partial Isocitrate lyase; ?	Reinscheid, D. J. et al. "Characterization of the isocitrate lyase gene from <i>Corynebacterium glutamicum</i> and biochemical analysis of the enzyme," J. Bacteriol., 176(12): 3474-3483 (1994)
X76875		ATPase beta-subunit	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," Antonie Van Leeuwenhoek, 64: 285-305 (1993)
X77034	tuf	Elongation factor Tu	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," Antonie Van Leeuwenhoek, 64: 285-305 (1993)

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X77384	recA		Billman-Jacobe, H. "Nucleotide sequence of a recA gene from <i>Corynebacterium glutamicum</i> ," DNA Seq., 4(6): 403-404 (1994)
X78491	aceB	Malate synthase	Reinscheid, D. J. et al. "Malate synthase from <i>Corynebacterium glutamicum</i> pta-ack operon encoding phosphotransacetylase: sequence analysis," Microbiology, 140: 3099-3108 (1994)
X80629	16S rDNA	16S ribosomal RNA	Rainey, F. A. et al. "Phylogenetic analysis of the genera <i>Rhodococcus</i> and <i>Nocardia</i> and evidence for the evolutionary origin of the genus <i>Nocardia</i> from within the radiation of <i>Rhodococcus</i> species," Microbiol., 141: 523-528 (1995)
X81191	gluA; gluB; gluC; gluD	Glutamate uptake system	Kronemeyer, W. et al. "Structure of the gluABCD cluster encoding the glutamate uptake system of <i>Corynebacterium glutamicum</i> ," J. Bacteriol., 177(5): 1152-1158 (1995)
X81379	dapE	Succinylaminopimelate desuccinylase	Wehrmann, A. et al. "Analysis of different DNA fragments of <i>Corynebacterium glutamicum</i> complementing dapE of <i>Escherichia coli</i> ," Microbiology, 40: 3349-56 (1994)
X82061	16S rDNA	16S ribosomal RNA	Ruimy, R. et al. "Phylogeny of the genus <i>Corynebacterium</i> deduced from analyses of small-subunit ribosomal DNA sequences," Int. J. Syst. Bacteriol., 45(4): 740-746 (1995)
X82928	asd; lysC	Aspartate-semialdehyde dehydrogenase; ?	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," J. Bacteriol., 177(24): 7255-7260 (1995)
X82929	proA	Gamma-glutamyl phosphate reductase	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," J. Bacteriol., 177(24): 7255-7260 (1995)
X84257	16S rDNA	16S ribosomal RNA	Pascual, C. et al. "Phylogenetic analysis of the genus <i>Corynebacterium</i> based on 16S rRNA gene sequences," Int. J. Syst. Bacteriol., 45(4): 724-728 (1995)
X85965	aroP; dapE	Aromatic amino acid permease; ?	Wehrmann, A. et al. "Functional analysis of sequences adjacent to dapE of <i>Corynebacterium glutamicum proline</i> reveals the presence of aroP, which encodes the aromatic amino acid transporter," J. Bacteriol., 177(20): 5991-5993 (1995)
X86157	argB; argC; argD; argF; argJ	Acetylglutamate kinase; N-acetyl-gamma-glutamyl-phosphate reductase; acetylornithine aminotransferase; ornithine carbamoyltransferase; glutamate N-acetyltransferase	Sakanyan, V. et al. "Genes and enzymes of the acetyl cycle of arginine biosynthesis in <i>Corynebacterium glutamicum</i> : enzyme evolution in the early steps of the arginine pathway," Microbiology, 142: 99-108 (1996)
X89084	pta; ackA	Phosphate acetyltransferase; acetate kinase	Reinscheid, D. J. et al. "Cloning, sequence analysis, expression and inactivation of the <i>Corynebacterium glutamicum</i> pta-ack operon encoding phosphotransacetylase and acetate kinase," Microbiology, 145: 503-513 (1999)
X89850	attB	Attachment site	Le Marrec, C. et al. "Genetic characterization of site-specific integration functions of phi AAU2 infecting "Arthrobacter aureus C70," J. Bacteriol., 178(7): 1996-2004 (1996)
X90356		Promoter fragment F1	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90357		Promoter fragment F2	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90358		Promoter fragment F10	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90359		Promoter fragment F13	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90360		Promoter fragment F22	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90361		Promoter fragment F34	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90362		Promoter fragment F37	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90363		Promoter fragment F45	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90364		Promoter fragment F64	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)

TABLE 2-continued

<u>GENES IDENTIFIED FROM GENBANK</u>			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
X90365		Promoter fragment F75	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142: 1297-1309 (1996)
X90366		Promoter fragment PF101	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142: 1297-1309 (1996)
X90367		Promoter fragment PF104	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142: 1297-1309 (1996)
X90368		Promoter fragment PF109	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," <i>Microbiology</i> , 142: 1297-1309 (1996)
X93513	amt	Ammonium transport system	Siewe, R. M. et al. "Functional and genetic characterization of the (methyl) ammonium uptake carrier of <i>Corynebacterium glutamicum</i> ," <i>J. Biol. Chem.</i> , 271(10): 5398-5403 (1996)
X93514	betP	Glycine betaine transport system	Peter, H. et al. "Isolation, characterization, and expression of the <i>Corynebacterium glutamicum</i> betP gene, encoding the transport system for the compatible solute glycine betaine," <i>J. Bacteriol.</i> , 178(17): 5229-5234 (1996)
X95649	orf4		Patek, M. et al. "Identification and transcriptional analysis of the dapB-ORF2-dapA-ORF4 operon of <i>Corynebacterium glutamicum</i> , encoding two enzymes involved in L-lysine synthesis," <i>Biotechnol. Lett.</i> , 19: 1113-1117 (1997)
X96471	lysE; lysG	Lysine exporter protein; Lysine export regulator protein	Vrljic, M. et al. "A new type of transporter with a new type of cellular function: L-lysine export from <i>Corynebacterium glutamicum</i> ," <i>Mol. Microbiol.</i> , 22(5): 815-826 (1996)
X96580	panB; panC; xylB	3-methyl-2-oxobutanoate hydroxymethyltransferase; pantoate-beta-alanine ligase; xylulokinase	Sahm, H. et al. "D-pantothenate synthesis in <i>Corynebacterium glutamicum</i> and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction," <i>Appl. Environ. Microbiol.</i> , 65(5): 1973-1979 (1999)
X96962		Insertion sequence IS1207 and transposase	Ramos, A. et al. "Cloning, sequencing and expression of the gene encoding elongation factor P in the amino-acid producer <i>Brevibacterium lactofermentum</i> (<i>Corynebacterium glutamicum</i> ATCC 13869)," <i>Gene</i> , 198: 217-222 (1997)
X99289		Elongation factor P	Mateos, L. M. et al. "Nucleotide sequence of the homoserine kinase (thrB) gene of the <i>Brevibacterium lactofermentum</i> ," <i>Nucleic Acids Res.</i> , 15(9): 3922 (1987)
Y00140	thrB	Homoserine kinase	Ishino, S. et al. "Nucleotide sequence of the meso-diaminopimelate D-dehydrogenase gene from <i>Corynebacterium glutamicum</i> ," <i>Nucleic Acids Res.</i> , 15(9): 3917 (1987)
Y00151	ddh	Meso-diaminopimelate D-dehydrogenase (EC 1.4.1.16)	Mateos, L. M. et al. "Nucleotide sequence of the homoserine dehydrogenase (thrA) gene of the <i>Brevibacterium lactofermentum</i> ," <i>Nucleic Acids Res.</i> , 15(24): 10598 (1987)
Y00476	thrA	Homoserine dehydrogenase	Peoples, O. P. et al. "Nucleotide sequence and fine structural analysis of the <i>Corynebacterium glutamicum</i> hom-thrB operon," <i>Mol. Microbiol.</i> , 2(1): 63-72 (1988)
Y00546	hom; thrB	Homoserine dehydrogenase; homoserine kinase	Honrubia, M. P. et al. "Identification, characterization, and chromosomal organization of the ftsZ gene from <i>Brevibacterium lactofermentum</i> ," <i>Mol. Gen. Genet.</i> , 259(1): 97-104 (1998)
Y08964	murC; ftsQ/divD; ftsZ	UPD-N-acetylmuramate-alanine ligase; division initiation protein or cell division protein; cell division protein	Peter, H. et al. "Isolation of the putP gene of <i>Corynebacterium glutamicum</i> and characterization of a low-affinity uptake system for compatible solutes," <i>Arch. Microbiol.</i> , 168(2): 143-151 (1997)
Y09163	putP	High affinity proline transport system	Peters-Wendisch, P. G. et al. "Pyruvate carboxylase from <i>Corynebacterium glutamicum</i> : characterization, expression and inactivation of the pyc gene," <i>Microbiology</i> , 144: 915-927 (1998)
Y09548	pyc	Pyruvate carboxylase	Patek, M. et al. "Analysis of the leuB gene from <i>Corynebacterium glutamicum</i> ," <i>Appl. Microbiol. Biotechnol.</i> , 50(1): 42-47 (1998)
Y09578	leuB	3-isopropylmalate dehydrogenase	Moreau, S. et al. "Site-specific integration of corynephage Phi-16: The construction of an integration vector," <i>Microbiol.</i> , 145: 539-548 (1999)
Y12472		Attachment site bacteriophage Phi-16	Peter, H. et al. " <i>Corynebacterium glutamicum</i> is equipped with four secondary carriers for compatible solutes: Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," <i>J. Bacteriol.</i> , 180(22): 6005-6012 (1998)
Y12537	proP	Proline/ectoine uptake system protein	Jakoby, M. et al. "Isolation of <i>Corynebacterium glutamicum</i> glnA gene encoding glutamine synthetase I," <i>FEMS Microbiol. Lett.</i> , 154(1): 81-88 (1997)
Y13221	glnA	Glutamine synthetase I	
Y16642	lpd	Dihydrolipoamide dehydrogenase	Moreau, S. et al. "Analysis of the integration functions of φ 304L: An integrase module among corynephages," <i>Virology</i> , 255(1): 150-159 (1999)
Y18059		Attachment site Corynephage 304L	Oguiza, J. A. et al. "A gene encoding arginyl-tRNA synthetase is located in the upstream region of the lysA gene in <i>Brevibacterium lactofermentum</i> : Regulation of argS-lysA cluster expression by arginine," <i>J. Bacteriol.</i> , 175(22): 7356-7362 (1993)
Z21501	argS; lysA	Arginyl-tRNA synthetase; diaminopimelate decarboxylase (partial)	

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
Z21502	dapA; dapB	Dihydrodipicolinate synthase; dihydrodipicolinate reductase	Pisabarro, A. et al. "A cluster of three genes (dapA, orf2, and dapB) of <i>Brevibacterium lactofermentum</i> encodes dihydrodipicolinate reductase, and a third polypeptide of unknown function," J. Bacteriol., 175(9): 2743-2749 (1993)
Z29563	thrC	Threonine synthase	Malumbres, M. et al. "Analysis and expression of the thrC gene of the encoded threonine synthase," Appl. Environ. Microbiol., 60(7):2209-2219 (1994)
Z46753	16S rDNA	Gene for 16S ribosomal RNA	Oguiza, J. A. et al "Multiple sigma factor genes in <i>Brevibacterium lactofermentum</i> : Characterization of sigA and sigB," J. Bacteriol., 178(2): 550-553 (1996)
Z49822	sigA	SigA sigma factor	Oguiza, J. A. et al "The galE gene encoding the UDP-galactose 4-epimerase of <i>Brevibacterium lactofermentum</i> is coupled transcriptionally to the dmdR gene," Gene, 177: 103-107 (1996)
Z49823	galE; dtxR	Catalytic activity UDP-galactose 4-epimerase; diphtheria toxin regulatory protein	Oguiza, J. A. et al "Multiple sigma factor genes in <i>Brevibacterium lactofermentum</i> : Characterization of sigA and sigB," J. Bacteriol., 178(2): 550-553 (1996)
Z49824	orf1; sigB	?; SigB sigma factor	Oguiza, J. A. et al "Cloning and characterization of an IS-like element present in the genome of <i>Brevibacterium lactofermentum</i> ATCC 13869," Gene, 170(1): 91-94 (1996)
Z66534		Transposase	Correia, A. et al. "Cloning and characterization of an IS-like element present in the genome of <i>Brevibacterium lactofermentum</i> ATCC 13869," Gene, 170(1): 91-94 (1996)

¹A sequence for this gene was published in the indicated reference. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

[0214]

TABLE 3

Corynebacterium and Brevibacterium Strains Which May be Used in the Practice of the Invention

Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21054							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19350							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19351							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19352							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19353							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19354							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19355							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19356							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21055							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21077							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21553							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21580							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	39101							
<i>Brevibacterium</i>	<i>butanicum</i>	21196							
<i>Brevibacterium</i>	<i>divaricatum</i>	21792	P928						
<i>Brevibacterium</i>	<i>flavum</i>	21474							
<i>Brevibacterium</i>	<i>flavum</i>	21129							
<i>Brevibacterium</i>	<i>flavum</i>	21518							
<i>Brevibacterium</i>	<i>flavum</i>				B11474				
<i>Brevibacterium</i>	<i>flavum</i>				B11472				
<i>Brevibacterium</i>	<i>flavum</i>	21127							
<i>Brevibacterium</i>	<i>flavum</i>	21128							
<i>Brevibacterium</i>	<i>flavum</i>	21427							
<i>Brevibacterium</i>	<i>flavum</i>	21475							
<i>Brevibacterium</i>	<i>flavum</i>	21517							
<i>Brevibacterium</i>	<i>flavum</i>	21528							
<i>Brevibacterium</i>	<i>flavum</i>	21529							
<i>Brevibacterium</i>	<i>flavum</i>				B11477				
<i>Brevibacterium</i>	<i>flavum</i>				B11478				
<i>Brevibacterium</i>	<i>flavum</i>	21127							
<i>Brevibacterium</i>	<i>flavum</i>				B11474				
<i>Brevibacterium</i>	<i>healii</i>	15527							
<i>Brevibacterium</i>	<i>ketoglutamicum</i>	21004							
<i>Brevibacterium</i>	<i>ketoglutamicum</i>	21089							
<i>Brevibacterium</i>	<i>ketosoreductum</i>	21914							

TABLE 3-continued

Corynebacterium and Brevibacterium Strains Which May be Used in the Practice of the Invention

Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
<i>Brevibacterium</i>	<i>lactofermentum</i>				70				
<i>Brevibacterium</i>	<i>lactofermentum</i>				74				
<i>Brevibacterium</i>	<i>lactofermentum</i>				77				
<i>Brevibacterium</i>	<i>lactofermentum</i>	21798							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21799							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21800							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21801							
<i>Brevibacterium</i>	<i>lactofermentum</i>			B11470					
<i>Brevibacterium</i>	<i>lactofermentum</i>			B11471					
<i>Brevibacterium</i>	<i>lactofermentum</i>	21086							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21420							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21086							
<i>Brevibacterium</i>	<i>lactofermentum</i>	31269							
<i>Brevibacterium</i>	<i>linens</i>	9174							
<i>Brevibacterium</i>	<i>linens</i>	19391							
<i>Brevibacterium</i>	<i>linens</i>	8377							
<i>Brevibacterium</i>	<i>paraffinolyticum</i>			11160					
<i>Brevibacterium</i>	spec.				717.73				
<i>Brevibacterium</i>	spec.				717.73				
<i>Brevibacterium</i>	spec.	14604							
<i>Brevibacterium</i>	spec.	21860							
<i>Brevibacterium</i>	spec.	21864							
<i>Brevibacterium</i>	spec.	21865							
<i>Brevibacterium</i>	spec.	21866							
<i>Brevibacterium</i>	spec.	19240							
<i>Corynebacterium</i>	<i>acetoacidophilum</i>	21476							
<i>Corynebacterium</i>	<i>acetoacidophilum</i>	13870							
<i>Corynebacterium</i>	<i>acetoglutamicum</i>		B11473						
<i>Corynebacterium</i>	<i>acetoglutamicum</i>		B11475						
<i>Corynebacterium</i>	<i>acetoglutamicum</i>	15806							
<i>Corynebacterium</i>	<i>acetoglutamicum</i>	21491							
<i>Corynebacterium</i>	<i>acetoglutamicum</i>	31270							
<i>Corynebacterium</i>	<i>acetophilum</i>		B3671						
<i>Corynebacterium</i>	<i>ammoniagenes</i>	6872				2399			
<i>Corynebacterium</i>	<i>ammoniagenes</i>	15511							
<i>Corynebacterium</i>	<i>fujikense</i>	21496							
<i>Corynebacterium</i>	<i>glutamicum</i>	14067							
<i>Corynebacterium</i>	<i>glutamicum</i>	39137							
<i>Corynebacterium</i>	<i>glutamicum</i>	21254							
<i>Corynebacterium</i>	<i>glutamicum</i>	21255							
<i>Corynebacterium</i>	<i>glutamicum</i>	31830							
<i>Corynebacterium</i>	<i>glutamicum</i>	13032							
<i>Corynebacterium</i>	<i>glutamicum</i>	14305							
<i>Corynebacterium</i>	<i>glutamicum</i>	15455							
<i>Corynebacterium</i>	<i>glutamicum</i>	13058							
<i>Corynebacterium</i>	<i>glutamicum</i>	13059							
<i>Corynebacterium</i>	<i>glutamicum</i>	13060							
<i>Corynebacterium</i>	<i>glutamicum</i>	21492							
<i>Corynebacterium</i>	<i>glutamicum</i>	21513							
<i>Corynebacterium</i>	<i>glutamicum</i>	21526							
<i>Corynebacterium</i>	<i>glutamicum</i>	21543							
<i>Corynebacterium</i>	<i>glutamicum</i>	13287							
<i>Corynebacterium</i>	<i>glutamicum</i>	21851							
<i>Corynebacterium</i>	<i>glutamicum</i>	21253							
<i>Corynebacterium</i>	<i>glutamicum</i>	21514							
<i>Corynebacterium</i>	<i>glutamicum</i>	21516							
<i>Corynebacterium</i>	<i>glutamicum</i>	21299							
<i>Corynebacterium</i>	<i>glutamicum</i>	21300							
<i>Corynebacterium</i>	<i>glutamicum</i>	39684							
<i>Corynebacterium</i>	<i>glutamicum</i>	21488							
<i>Corynebacterium</i>	<i>glutamicum</i>	21649							
<i>Corynebacterium</i>	<i>glutamicum</i>	21650							
<i>Corynebacterium</i>	<i>glutamicum</i>	19223							
<i>Corynebacterium</i>	<i>glutamicum</i>	13869							
<i>Corynebacterium</i>	<i>glutamicum</i>	21157							
<i>Corynebacterium</i>	<i>glutamicum</i>	21158							
<i>Corynebacterium</i>	<i>glutamicum</i>	21159							
<i>Corynebacterium</i>	<i>glutamicum</i>	21355							

TABLE 3-continued

<u>Corynebacterium and Brevibacterium Strains Which May be Used in the Practice of the Invention</u>									
Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
<i>Corynebacterium</i>	<i>glutamicum</i>	31808							
<i>Corynebacterium</i>	<i>glutamicum</i>	21674							
<i>Corynebacterium</i>	<i>glutamicum</i>	21562							
<i>Corynebacterium</i>	<i>glutamicum</i>	21563							
<i>Corynebacterium</i>	<i>glutamicum</i>	21564							
<i>Corynebacterium</i>	<i>glutamicum</i>	21565							
<i>Corynebacterium</i>	<i>glutamicum</i>	21566							
<i>Corynebacterium</i>	<i>glutamicum</i>	21567							
<i>Corynebacterium</i>	<i>glutamicum</i>	21568							
<i>Corynebacterium</i>	<i>glutamicum</i>	21569							
<i>Corynebacterium</i>	<i>glutamicum</i>	21570							
<i>Corynebacterium</i>	<i>glutamicum</i>	21571							
<i>Corynebacterium</i>	<i>glutamicum</i>	21572							
<i>Corynebacterium</i>	<i>glutamicum</i>	21573							
<i>Corynebacterium</i>	<i>glutamicum</i>	21579							
<i>Corynebacterium</i>	<i>glutamicum</i>	19049							
<i>Corynebacterium</i>	<i>glutamicum</i>	19050							
<i>Corynebacterium</i>	<i>glutamicum</i>	19051							
<i>Corynebacterium</i>	<i>glutamicum</i>	19052							
<i>Corynebacterium</i>	<i>glutamicum</i>	19053							
<i>Corynebacterium</i>	<i>glutamicum</i>	19054							
<i>Corynebacterium</i>	<i>glutamicum</i>	19055							
<i>Corynebacterium</i>	<i>glutamicum</i>	19056							
<i>Corynebacterium</i>	<i>glutamicum</i>	19057							
<i>Corynebacterium</i>	<i>glutamicum</i>	19058							
<i>Corynebacterium</i>	<i>glutamicum</i>	19059							
<i>Corynebacterium</i>	<i>glutamicum</i>	19060							
<i>Corynebacterium</i>	<i>glutamicum</i>	19185							
<i>Corynebacterium</i>	<i>glutamicum</i>	13286							
<i>Corynebacterium</i>	<i>glutamicum</i>	21515							
<i>Corynebacterium</i>	<i>glutamicum</i>	21527							
<i>Corynebacterium</i>	<i>glutamicum</i>	21544							
<i>Corynebacterium</i>	<i>glutamicum</i>	21492							
<i>Corynebacterium</i>	<i>glutamicum</i>		B8183						
<i>Corynebacterium</i>	<i>glutamicum</i>		B8182						
<i>Corynebacterium</i>	<i>glutamicum</i>		B12416						
<i>Corynebacterium</i>	<i>glutamicum</i>		B12417						
<i>Corynebacterium</i>	<i>glutamicum</i>		B12418						
<i>Corynebacterium</i>	<i>glutamicum</i>		B11476						
<i>Corynebacterium</i>	<i>glutamicum</i>	21608							
<i>Corynebacterium</i>	<i>lilium</i>		P973						
<i>Corynebacterium</i>	<i>nitrophilus</i>	21419			11594				
<i>Corynebacterium</i>	spec.		P4445						
<i>Corynebacterium</i>	spec.		P4446						
<i>Corynebacterium</i>	spec.	31088							
<i>Corynebacterium</i>	spec.	31089							
<i>Corynebacterium</i>	spec.	31090							
<i>Corynebacterium</i>	spec.	31090							
<i>Corynebacterium</i>	spec.	31090							
<i>Corynebacterium</i>	spec.	15954			20145				
<i>Corynebacterium</i>	spec.	21857							
<i>Corynebacterium</i>	spec.	21862							
<i>Corynebacterium</i>	spec.	21863							

ATCC: American Type Culture Collection, Rockville, MD, USA

FERM: Fermentation Research Institute, Chiba, Japan

NRRL: ARS Culture Collection, Northern Regional Research Laboratory, Peoria, IL, USA

CECT: Colección Espanola de Cultivos Tipo, Valencia, Spain

NCIMB: National Collection of Industrial and Marine Bacteria Ltd., Aberdeen, UK

CBS: Centraalbureau voor Schimmelcultures, Baarn, NL

NCTC: National Collection of Type Cultures, London, UK

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany

For reference see Sugawara, H. et al. (1993) World directory of collections of cultures of microorganisms: Bacteria, fungi and yeasts (4th edn), World federation for culture collections world data center on microorganisms, Saimata, Japan.

TABLE 4
ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit (NT)	Length	Accession	Name of Genbank Hit	% homology (GAP)	Source of Genbank Hit	Date of Deposit
rx00009	1023	GB_IN1: CELZK563	29655	U40061	<i>Caenorhabditis elegans</i> cosmid ZK563.	33,694 9-Nov-95	<i>Caenorhabditis elegans</i>	36,040 9-Nov-95
		GB_IN1: CELZK563	29655	U40061	<i>Caenorhabditis elegans</i> cosmid ZK563.	38,442 19-Jun-98	<i>Mycobacterium tuberculosis</i>	63,158 06-MAR-1997
rx00010	810	GB_BAI: MTICY164	39150	295150	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 135/162.	38,375 16-Aug-96	<i>Streptomyces hygroscopicus</i>	36,217 12-Jan-99
		GB_BAI: MTFTSX	4068	X70031	S. hygroscopicus gene cluster for polyketide immunosuppressant rapamycin.	36,217 12-Jan-99	<i>Caenorhabditis elegans</i>	36,217 12-Jan-99
rx00024	1068	GB_BAI: SHGCP1R	107379	X86780	<i>Caenorhabditis elegans</i> chromosome V clone Y113G7, ** SEQUENCING IN PROGRESS, **, in unordered pieces.	35,824 20-Sep-99	<i>Arabidopsis thaliana</i>	35,824 20-Sep-99
		GB_HTG1: CEY113G7_31	10000	AL031113	<i>Caenorhabditis elegans</i> chromosome V clone Y113G7, ** SEQUENCING IN PROGRESS, **, in unordered pieces.	38,826 3-Aug-99	<i>Homo sapiens</i>	38,826 3-Aug-99
		GB_HTG1: CEY113G7_31	10000	AL031113	<i>Caenorhabditis elegans</i> chromosome V clone Y113G7, ** SEQUENCING IN PROGRESS, **, in unordered pieces.	38,826 3-Aug-99	<i>Borderella pertussis</i>	38,826 3-Aug-99
		GB_PL2: ATFC12	111945	AL022224	<i>Arabidopsis thaliana</i> DNA chromosome 4, BAC clone F1C12 (ESSA project).	37,379 3-Aug-99	<i>Borderella pertussis</i>	45,756 17-Apr-95
rx00048	1782	GB_HTG3: AC008905	129915	AC008905	<i>Homo sapiens</i> chromosome 5 clone C1TB-H1_2259114, ** SEQUENCING IN PROGRESS, **, 40 unordered pieces.	47,119 14-Jan-95	<i>Borderella pertussis</i>	45,756 17-Apr-95
		GB_HTG3: AC008905	129915	AC008905	<i>Homo sapiens</i> chromosome 5 clone C1TB-H1_2259114, ** SEQUENCING IN PROGRESS, **, 40 unordered pieces.	45,756 17-Apr-95	<i>Borderella pertussis</i>	45,756 17-Apr-95
		GB_HTG3: AC008905	129915	AC008905	<i>Homo sapiens</i> chromosome 5 clone C1TB-H1_2259114, ** SEQUENCING IN PROGRESS, **, 40 unordered pieces.	45,756 17-Apr-95	<i>Homo sapiens</i>	39,010 3-Aug-95
rx00070	555	GB_BA2: BPEFUR	1003	L31851	<i>Borderella pertussis</i> DNA repair protein (recN) gene, partial cds; iron regulatory protein (fir) gene, complete cds.	40,661 23-Feb-99	<i>Candida albicans</i>	36,222 10-Nov-98
		GB_BA2: BPU11699	537	U11699	<i>Bordetella pertussis</i> ferric uptake regulator (fir) gene, complete cds.	35,445 1-Nov-96	<i>Homo sapiens</i>	33,050 17-Aug-99
		GB_BAI: BTUFURGEN	1106	248227	<i>B. pertussis</i> fir gene for ferric uptake regulator and partial recN gene.	34,439 1-Nov-96	<i>Mycobacterium leprae</i>	34,439 1-Nov-96
rx00078	537	GB_PRS: HUMCOL2A1Z3	1001	L10347	Human pro-alpha1 type II collagen (COL2A1) gene exons 1-54, complete cds.	34,439 1-Nov-96	<i>Mycobacterium tuberculosis</i>	38,377 09-MAR-1995
		GB_HTG2: AC006721	135550	AC006721	<i>Caenorhabditis elegans</i> clone Y18H1, ** SEQUENCING IN PROGRESS, **, 5 unordered pieces.	34,439 1-Nov-96	<i>Mycobacterium tuberculosis</i>	61,575 22-MAR-1996
		GB_HTG2: AC006721	135550	AC006721	<i>Caenorhabditis elegans</i> clone Y18H1, ** SEQUENCING IN PROGRESS, **, 5 unordered pieces.	34,439 1-Nov-96	<i>Mycobacterium tuberculosis</i>	37,773 01-DEC-1998
		GB_RO: MMCGT16	3009	U48896	<i>Mus musculus</i> UDP-galactose: ceramide galactosyltransferase (Cgt) gene, exon 6 and complete cds.	34,439 1-Nov-96	<i>Mycobacterium tuberculosis</i>	36,984 18-Jun-98
		GB_RO: MMCGT6	3009	U48896	<i>Mus musculus</i> UDP-galactose: ceramide galactosyltransferase (Cgt) gene, exon 6 and complete cds.	34,439 1-Nov-96	<i>Mycobacterium tuberculosis</i>	38,551 19-MAY-1999
rx00088	899	GB_HTG2: AC006721	38874	AL035501	<i>C. albicans</i> cosmid Ca4IC10.	34,439 1-Nov-96	<i>Candida albicans</i>	34,439 1-Nov-96
		GB_PR4: AC007115	180821	AC007115	<i>Homo sapiens</i> chromosome 12 clone 91705, complete sequence.	34,439 17-Aug-99	<i>Homo sapiens</i>	34,439 17-Aug-99
		GB_PR4: AC007115	180821	AC007115	<i>Homo sapiens</i> chromosome 12 clone 91705, complete sequence.	34,439 17-Aug-99	<i>Mycobacterium tuberculosis</i>	60,639 17-Jun-98
rx00135	1377	GB_BAI: MTICY373	35516	273419	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 5/162.	38,377 09-MAR-1995	<i>Mycobacterium leprae</i>	38,377 09-MAR-1995
		GB_BAI: MLU15186	36241	U15186	<i>Mycobacterium leprae</i> cosmid L471.	38,377 09-MAR-1995	<i>Mycobacterium tuberculosis</i>	38,377 09-MAR-1995
		GB_BAI: MTMURAGEN	1257	X86711	<i>M. tuberculosis</i> mura gene.	38,377 09-MAR-1995	<i>Mycobacterium tuberculosis</i>	38,377 09-MAR-1995
rx00143	1605	GB_PAT: 192051	1107	I92051	Sequence 18 from Patent US 5726299.	38,377 09-MAR-1995	Unknown.	37,773 01-DEC-1998
		GB_PAT: 178761	1107	I78761	Sequence 17 from Patent US 5693781.	38,377 09-MAR-1995	<i>Mycobacterium tuberculosis</i>	37,773 3-Apr-98
		GB_BAI: MTICY28	40163	295890	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 79/162.	38,377 09-MAR-1995	<i>Homo sapiens</i>	36,984 18-Jun-98
rx00177	1191	GB_GSS14: AQ543786	345	AQ543786	RPCI-11-3651L6-TV RPCI-11 <i>Homo sapiens</i> genomic clone RPCI-11-3651L6, genomic survey sequence.	38,377 09-MAR-1995		

TABLE 4-continued

ALIGNMENT RESULTS

ID #	Length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
		GB_P12: AF017646	3394	AF017646	<i>Schizosaccharomyces pombe</i> TFIH subunit p47 (tfh47) gene, complete cds.	<i>Schizosaccharomyces pombe</i>	38,122 17-MAR-1999	
		GB_PLI: SPCC1682	37404	AL031525	S. pombe chromosome III cosmid c1682.	<i>Schizosaccharomyces pombe</i>	33,983 14-DEC-1998	
rxa00178	1008	GB_BA1: AB016258	2260	AB016258	<i>Arthrobacter</i> sp. gene for maleylacetate reductase and hydroxyquinol 1,2-dioxygenase, partial and complete cds.	<i>Arthrobacter</i> sp.	65,182 8-Sep-99	
		GB_BA1: CGPUTP	3791	Y09163	<i>C. glutamicum</i> putP gene.	<i>Corynebacterium glutamicum</i>	38,806 8-Sep-97	
		GB_STS: G05495	271	G05495	Human STS W1-5918.	<i>Homo sapiens</i>	39,925 8-Jun-95	
		GB_BA1: MTCY22G10	35420	284724	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 21/162.	<i>Mycobacterium tuberculosis</i>	39,976 17-Jun-98	
rxa00277	1684	GB_IN1: CELT03F1	38643	U88169	<i>Caenorhabditis elegans</i> cosmid T03F1.	<i>Caenorhabditis elegans</i>	35,127 7-Feb-97	
		GB_IN2: CELK02A2	38261	U23171	<i>Caenorhabditis elegans</i> cosmid K02A2.	<i>Caenorhabditis elegans</i>	36,166 21-MAY-1999	
rxa00372	1200	GB_IN2: AC005452	79333	AC005452	<i>Drosophila melanogaster</i> , chromosome 2R, region 43B2-4C2, P1 clone DS07185, complete sequence.	<i>Drosophila melanogaster</i>	37,006 26-Nov-98	
		GB_IN2: AC005452	79333	AC005452	<i>Drosophila melanogaster</i> , chromosome 2R, region 43B2-4C2, P1 clone DS07185, complete sequence.	<i>Drosophila melanogaster</i>	34,907 26-Nov-98	
		GB_IN1: CELW03F8	34766	AF039041	<i>Caenorhabditis elegans</i> cosmid W03F8.	<i>Caenorhabditis elegans</i>	40,712 1-Jan-98	
rxa00389	1683	GB_IN1: AB010703	772	AB010703	<i>Theileria</i> sp. gene for major piroplasm surface protein, partial cds, isolate Kamphaeng Saen.	<i>Theileria</i> sp.	40,285 18-Apr-98	
		GB_BA1: LLU08911	619	U08911	<i>Lactobacillus leichmannii</i> putative D-alanine ligase (ddl) gene, partial cds.	<i>Lactobacillus leichmannii</i>	40,194 16-Feb-96	
		GB_IN1: TPMS1	822	248740	<i>T. parva</i> Tpm1 gene for merozoite surface glycoprotein.	<i>Theileria parva</i>	38,902 15-MAY-1995	
rxa00467	792	GB_PR4: DJ293M10	202267	AF111167	<i>Homo sapiens</i> jun dimerization protein gene, partial cds; cfos gene, complete cds; and unknown gene.	<i>Homo sapiens</i>	37,995 7-Apr-99	
		GB_PR4: DJ293M10	202267	AF111167	<i>Homo sapiens</i> jun dimerization protein gene, partial cds; cfos gene, complete cds; and unknown gene.	<i>Homo sapiens</i>	36,639 7-Apr-99	
		GB_IN1: CEW01C9	21493	ZA9969	<i>Caenorhabditis elegans</i> cosmid W01C9, complete sequence.	<i>Caenorhabditis elegans</i>	37,980 23-Nov-98	
		GB_PR4: AC007206	42732	AC007206	<i>Homo sapiens</i> chromosome 19, cosmid R27370, complete sequence.	<i>Homo sapiens</i>	34,982 4-Apr-99	
		GB_EST26: AI344735	462	A1344735	cp05a10.x1 NCL_CGAP_Kid5 <i>Homo sapiens</i> cDNA clone IMAGE: 1917114 3' similar to gb: M15800 T-LYMPHOCYTE MATURATION-ASSOCIATED PROTEIN (HUMAN), mRNA sequence.	<i>Homo sapiens</i>	42,675 2-Feb-99	
rxa00499	1404	GB_PR4: AC006479	161837	AC006479	<i>Homo sapiens</i> clone D1051J04, complete sequence.	<i>Homo sapiens</i>	38,462 11-Nov-99	
		GB_HTG2: AC007111	84245	AC007111	<i>Homo sapiens</i> chromosome 16 clone 1-8F, *** SEQUENCING IN PROGRESS *** , 2 ordered pieces.	<i>Homo sapiens</i>	37,931 18-MAR-1999	
rxa00508	1206	GB_PR4: AC006479	161837	AC006479	<i>Homo sapiens</i> chromosome 16 clone 1-8F, *** SEQUENCING IN PROGRESS *** , 2 ordered pieces.	<i>Homo sapiens</i>	39,401 7-Jul-99	
		GB_HTG2: AC007111	84245	AC007111	<i>Homo sapiens</i> chromosome 16 clone 1-8F, *** SEQUENCING IN PROGRESS *** , 2 ordered pieces.	<i>Homo sapiens</i>	41,244 2-Apr-96	
rxa00569	1149	GB_VI: AF141890	1791	AF141890	Columbid herpesvirus 1 DNA-dependent DNA polymerase gene, partial cds. Sequence 1 from patent US 5460951.	columbid herpesvirus 1 Unknown.	41,244 29-Sep-97	
		GB_PAT: II5213	3728	II5213	Sequence 1 from patent US 5460951.	<i>Mus</i> sp.	34,148 12-DEC-1997	
		GB_PAT: E07353	3728	E07353	cDNA encoding bone-related carboxypeptidase-like protein, OSF-5.	<i>Caenorhabditis elegans</i>		
		GB_HTG1: CEFY70G10	152184	AL020987	<i>Caenorhabditis elegans</i> chromosome III clone Y70G10, *** SEQUENCING IN PROGRESS *** , in unordered pieces.	<i>Caenorhabditis elegans</i>		
rxa00612	1077	GB_HTG2: AC005020	177756	AC005020	<i>Homo sapiens</i> clone GS259H13, *** SEQUENCING IN PROGRESS *** , 4 unordered pieces.	<i>Homo sapiens</i>	34,551 12-Jun-98	
		GB_HTG2: AC005020	177756	AC005020	<i>Homo sapiens</i> clone GS259H13, *** SEQUENCING IN PROGRESS *** , 4 unordered pieces.	<i>Homo sapiens</i>	34,551 12-Jun-98	

TABLE 4-continued

ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa00615	705	GB_GSS15_AQ622921	177756 AC005020	<i>Homo sapiens</i> clone GS259H13, *** SEQUENCING IN PROGRESS *** , 4 unordered pieces, HS_5351_A1_A08_ITA RPLC-11 Human Male BAC Library/ <i>Homo sapiens</i> genomic clone Plate = 927 Col = 15 Row = A, genomic survey sequence.	<i>Homo sapiens</i>	37,628 12-Jun-98	38,234 16-Jun-99
		GB_GSS3: B36703	517 AQ022921	HS_1041_B1-B12-MR.abi CITE Human Genomic Sperm Library <i>Chomo sapiens</i> genomic clone Plate = CT 823 Col = 23 Row = D, genomic sequence.	<i>Homo sapiens</i>	44,981 17-OCT-1997	
		GB_EST25: AI245926	432 B36703	sequence. qk33c08_x1 NCL_CGAP_CoB <i>Homo sapiens</i> cDNA clone IMAGE: 1870766 3' similar to SW: COPG_BOVIN P53620 COATOMER GAMMA SUBUNIT,	<i>Homo sapiens</i>	38,902 28-Jan-99	
rxa00621	906	GB_EST1: D36491	572 AI245926	CDNA clone qk33c08_x1 NCL_CGAP_CoB <i>Homo sapiens</i> cDNA clone IMAGE: 1870766 mRNA sequence. CELK033GYF Yuji Kohara unpublished cDNA <i>Caenorhabditis elegans</i> CDNA clone NH0311103, *** SEQUENCING IN PROGRESS *** , 3 unordered pieces.	<i>Caenorhabditis elegans</i>	40,390 8-Aug-94	35,477 18-MAY-1999
		GB_IN2: CELC16A3	360 D36491	<i>Caenorhabditis elegans</i> cosmid C16A3. <i>Homo sapiens</i> clone NH0311103, *** SEQUENCING IN PROGRESS *** , 3 unordered pieces.	<i>Homo sapiens</i>	38,636 13-Aug-99	38,636 13-Aug-99
rxa00622	1539	GB_BAI: AB004795	3039 AB004795	<i>Pseudomonas</i> sp. gene for dipeptidyl aminopeptidase, complete cds. <i>Moraxella lacunata</i> gene for protease II, complete cds.	<i>Pseudomonas</i> sp.	54,721 5-Feb-99	50,167 8-Feb-99
		GB_BA1: MBOPII	2392 D38405	<i>Trypanosoma brucei</i> oligopeptidase B (opb) gene, complete cds.	<i>Moraxella lacunata</i>	48,076 08-OCT-1999	48,076 08-OCT-1999
		GB_IN2: AF078916	2960 AF078916	<i>Rhodococcus rhodochrous</i> catchetol 1,2-dioxygenase (catA) gene, complete cds.	<i>Trypanosoma brucei</i>	66,940 27-Aug-98	<i>Rhodococcus rhodochrous</i>
rxa00639	978	GB_BA2: AF043741	1223 AF043741	<i>Rhodococcus erythropolis</i> DNA for catechol 1,2-dioxygenase, complete cds.	<i>Rhodococcus erythropolis</i>	65,440 1-Sep-99	63,617 24-Sep-97
		GB_BA1: D83237	1626 DS3237	<i>Rhodococcus opacus</i> catR, catA, catC genes and five ORFs.	<i>Rhodococcus opacus</i>	59,863 20-MAY-1999	<i>Pseudomonas putida</i>
		GB_BA1: ROX9622	7224 X99622	<i>Pseudomonas putida</i> plasmid pDK1 toluate 1,2 dioxygenase subunit (xylX), toluate 1,2 dioxygenase subunit (xylY), and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxyoctahexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	<i>Pseudomonas putida</i>		
rxa00641	1614	GB_BA2: AF134348	5000 AF134348	<i>Pseudomonas putida</i> plasmid pWW0 toluate 1,2 dioxygenase subunit (xylX), toluate 1,2 dioxygenase subunit (xylY), and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxyoctahexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	<i>Pseudomonas putida</i>		
		GB_BA1: PWWXYL	9037 M64747	<i>Pseudomonas putida</i> plasmid pWW0 meta operon, 5' genes.	Plasmid pWW0	59,588 26-Apr-93	55,410 3-Apr-97
		GB_BA1: FCCBDABC	3548 X79076	<i>Burkholderia cepacia</i>			
rxa00642	615	GB_BA2: AF134348	5000 AF134348	<i>Pseudomonas putida</i> plasmid pWW0 meta operon, 5' genes.	<i>Pseudomonas putida</i>	60,920 20-MAY-1999	
		GB_BA1: PWWXYL	9037 M64747	<i>Oryza sativa</i>			
		GB_GSS11: AQ274007	637 AQ274007	toluate 1,2 dioxygenase subunit (xylY) and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxyoctahexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	<i>Oryza sativa</i>		
rxa00643	1659	GB_BA2: AF134348	5000 AF134348	<i>Pseudomonas putida</i> plasmid pDK1 toluate 1,2 dioxygenase subunit (xylX), toluate 1,2 dioxygenase subunit (xylY) and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxyoctahexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	<i>Pseudomonas putida</i>	53,871 20-MAY-1999	
		GB_BA1: PWWXYL	9037 M64747	<i>Pseudomonas putida</i> plasmid pWW0 meta operon, 5' genes.	Plasmid pWW0	58,756 26-Apr-93	41,390 3-Nov-98
		GB_EST22: A1020666	328 A1020666	toluate 1,2 dioxygenase subunit (xylY) and toluate 1,2 dioxygenase subunit (xylZ) genes, complete cds; and 1,2-dihydroxyoctahexa-3,5-diene carboxylate dehydrogenase (xylL) gene, partial cds.	<i>Mus musculus</i>	52,603 26-Apr-93	43,865 16-Jun-98
rxa00644	951	GB_BA1: PWWXYL	9037 M64747	<i>Pseudomonas putida</i> plasmid pWW0 meta operon, 5' genes.	Plasmid pWW0	55,626 26-Apr-93	50,410 20-MAY-1999
		GB_BA2: AF134348	5000 AF134348	<i>Pseudomonas putida</i>			

TABLE 4-continued

ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
					toluate 1,2 dioxygenase subunit (xyY), and toluate 1,2 dioxygenase subunit (xyY2) genes, complete cds, and 1,2-dihydroxycyclohexa-3,5-diene carboxylate dehydrogenase (XYL) gene, partial cds.			40,138 28-Aug-98
GB_EST22; AI038396	438	AI038396	438		ox21g_0x1_Sores_fetal_liver_spleen_1NFLS_SI Homo sapiens cDNA sequence.	Homo sapiens		
GB_EST16; C26090	816	C26090	414	C26090	Rice callus cDNA Oryza sativa cDNA clone C1617_1A, mRNA sequence.	Oryza sativa		40,636 6-Aug-97
GB_BAI; MTV017	1497	AL021897	67200	AL021897	Mycobacterium tuberculosis H37Rv complete genome; segment 48/162.	Mycobacterium tuberculosis	57,976 24-Jun-99	
GB_BAI; MLCB1222		34714 AL049491			Mycobacterium leprae cosmid B1222.	Mycobacterium leprae	39,669 27-Aug-99	
GB_HTG2; AC007482		155357 AC007482			Homo sapiens clone hRPK_56_A_1, *** SEQUENCING IN PROGRESS ***, Homo sapiens 6 unordered pieces.		36,154 05-MAY-1999	
GB_BAI; SC3C8	915	AL023861	33095	AL023861	Streptomyces coelicolor cosmid 3C8.	Streptomyces coelicolor	36,836 15-Jan-99	
GB_PR3; AC005736		215441 AC005736			Homo sapiens chromosome 16, BAC clone 462G18 (LANI), complete sequence.	Homo sapiens	42,027 01-OCT-1998	
GB_IN2; AC005719		188357 AC005719			Drosophila melanogaster, chromosome 2L, region 3RA5-3RB4, BAC clone BACR48M05, complete sequence.	Drosophila melanogaster	35,531 27-OCT-1999	
GB_HTG2; HS1473116	999	AL109942	203460	AL109942	Homo sapiens chromosome 6 clone RP3-473J16 map q25.3-26, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	37,295 03-DEC-1999	
GB_HTG2; HS1473116		203460 AL109942			Homo sapiens chromosome 6 clone RP3-473J16 map q25.3-26, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	Homo sapiens	37,295 03-DEC-1999	
GB_PR2; HSU91327		129252 U91327			Human chromosome 12p15 BAC clone CTR875SK-99D8 complete sequence.	Homo sapiens	35,650 21-Aug-97	
GB_BA2; AF010184	1629	AF010184	1494		Pseudomonas aeruginosa coenzyme A transferase PsecoA (psecoA) gene, complete cds.	Pseudomonas aeruginosa	56,472 18-Jul-98	
GB_PAT; I92043		713 I92043			Sequence 10 from patent US 5726299.	Unknown.	92,701 01-DEC-1998	
GB_PAT; I78754		713 I78754			Sequence 10 from patent US 5693781.	Unknown.	92,701 3-Apr-98	
GB_BAI; MTPST2GN	1248	GB_BAI; D90907	1347 ZA8056	D90907	M. tuberculosis PsS-2 gene.	Mycobacterium tuberculosis	47,791 24-Apr-99	
GB_BAI; D90907		132419 D90907			Synechocystis sp. PCC6803 complete genome, 9/27, 1056467-1188885.	Synechocystis sp.	35,536 7-Feb-99	
GB_PLI; SCX11RA	2025	36849 X91258			Synechocystis sp. PCC6803 complete genome, 9/27, 1056467-1188885.	Synechocystis sp.	38,006 7-Feb-99	
GB_PL2; YSCL9606		29154 U53881			S. cerevisiae DNA from chromosome XII right arm including ACE2, CK1, PDC5, SLS1, PUT1 and tRNA-Asp genes.	Saccharomyces cerevisiae	36,122 13-OCT-1995	
GB_PLI; SCX11RA		36849 X91258			S. cerevisiae DNA from chromosome XII cosmid 9606.	Saccharomyces cerevisiae	36,122 25-OCT-1997	
GB_PR4; AC004841	1320	AC004841	132072		PDC5, SLS1, PUT1 and tRNA-Asp genes.	Saccharomyces cerevisiae	37,198 13-OCT-1995	
GB_HTG2; AC006706		180664 AC006706			Homo sapiens PAC clone D10607123 from 7q21.2-q31.1, complete sequence.	Homo sapiens	37,452 18-MAR-1999	
GB_HTG2; AC006706		180664 AC006706			Caenorhabditis elegans clone Y110A2, *** SEQUENCING IN PROGRESS *** , 4 unordered pieces.	Caenorhabditis elegans	34,824 23-Feb-99	
GB_HTG2; AC006706		180664 AC006706			Caenorhabditis elegans clone Y110A2, *** SEQUENCING IN PROGRESS *** , 4 unordered pieces.	Caenorhabditis elegans	34,824 23-Feb-99	

TABLE 4-continued
ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa00857	1313	GB_BA1: MTV002	56414	AL008967	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 122/162.	<i>Mycobacterium tuberculosis</i>	38,080 17-Jun-98	
		GB_BA1: MSGY154	40221	AD000002	<i>Mycobacterium tuberculosis</i> sequence from clone y154.	<i>Mycobacterium tuberculosis</i>	68,345 03-DEC-1996	
		GB_BA1: MLCB33	42224	294723	<i>Mycobacterium leprae</i> cosmid B33.	<i>Mycobacterium leprae</i>	38,824 24-Jun-97	
rxa00877	1788	GB_PAT: 192050	567	192050	Sequence 17 from patent US 5726299.	Unknown.	62,787 01-DEC-1998	
		GB_PAT: 178760	567	178760	Sequence 16 from patent US 5693781.	Unknown.	62,787 3-Apr-98	
rxa00888	1140	GB_BA2: AE000426	10240	AE000426	<i>Escherichia coli</i> K-12 MG1655 section 316 of 400 of the complete genome.	<i>Escherichia coli</i>	36,456 12-Nov-98	
		GB_BA1: MTICY27	27548	295208	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 104/162.	<i>Mycobacterium tuberculosis</i>	40,165 17-Jun-98	
		GB_BA1: U100016	42931	U00016	<i>Mycobacterium leprae</i> cosmid B1937.	<i>Mycobacterium leprae</i>	58,444 01-MAR-1994	
		GB_BA2: ECU82598	136742	U82598	<i>Escherichia coli</i> genomic sequence of minutes 9 to 12.	<i>Escherichia coli</i>	37,876 15-Jan-97	
rxa00892	1017	GB_BA2: AE000817	13157	AE000817	<i>Methanobacterium thermoautotrophicum</i> from bases 251486 to 264642 (section 23 of 148) of the complete genome.	<i>Methanobacterium</i>	36,710 15-Nov-97	
		GB_EST29: AI620549	239	AI620549	tw95b07.x1 NCL_CGAP_Gs4/Homo sapiens cDNA clone IMAGE: 2258773 3' similar to gb: X60708_mal DIPEPTIDYL PEPTIDASE IV (HUMAN); mRNA sequence.	<i>Homo sapiens</i>	38,075 21-Apr-99	
		GB_BA2: AE000817	13157	AE000817	<i>Methanobacterium thermoautotrophicum</i> from bases 251486 to 264642 (section 23 of 148) of the complete genome.	<i>Methanobacterium</i>	35,630 15-Nov-97	
rxa00897	1128	GB_PR3: HS246D7	28011	AL031843	Human DNA sequence from clone 246D7 on chromosome 22q13.1-13.33. Contains ESTs, a GSS and an STS, complete sequence.	<i>Homo sapiens</i>	38,724 23-Nov-99	
		GB_PR3: HSD1185D5	24387	AL118498	Human DNA sequence from clone 185D5 on chromosome 22, complete sequence.	<i>Homo sapiens</i>	37,021 23-Nov-99	
		GB_PR3: HS246D7	28011	AL031843	Human DNA sequence from clone 246D7 on chromosome 22q13.1-13.33. Contains ESTs, a GSS and an STS, complete sequence.	<i>Homo sapiens</i>	36,054 23-Nov-99	
rxa00944	1095	GB_BA1: ECU68759	1531	U68759	<i>Enterobacter cloacae</i> penicillityrihol tetranitrate reductase (nor) gene, complete cds.	<i>Enterobacter cloacae</i>	43,041 14-DEC-1996	
		GB_PAT: A59288	1531	A59288	Sequence 1 from Patent WO9703201.	unidentified	43,041 06-MAR-1998	
		GB_EST23: AI099394	601	AI099394	Sequence 1 from Patent WO9703201. Mus musculus cDNA clone ue32a9 y1 Sugano mouse liver mila. <i>Mus musculus</i> c-mer tyrosine kinase receptor mRNA, complete (MOUSE), mRNA sequence.	<i>Mus musculus</i>	37,225 20-Aug-98	
rxa00964	1248	GB_HTG6: AC009794	152794	AC009794	<i>Homo sapiens</i> chromosome 4 clone RP11-343C10 map 4, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.	<i>Homo sapiens</i>	34,762 03-DEC-1999	
		GB_HTG6: AC009794	152794	AC009794	<i>Homo sapiens</i> chromosome 4 clone RP11-343C10 map 4, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.	<i>Homo sapiens</i>	35,708 03-DEC-1999	
rxa00982	1629	GB_BA1: BLARGS	2501	211501	<i>B. lacfermentum</i> argS and lysA genes for arginyl-tRNA synthetase and diaminopimelate decarboxylase (partial).	<i>Corynebacterium glutamicum</i>	39,003 28-DEC-1993	
		GB_BA1: CGXLYSA	2344	X54740	<i>Corynebacterium glutamicum</i> argS-lysA operon gene for the upstream region of the arginyl-tRNA synthetase and diaminopimelate decarboxylase (EC 4.1.1.20).	<i>Corynebacterium glutamicum</i>	41,435 30-Jun-93	
		GB_PAT: E14508	3579	E14508	DNA encoding Brevibacterium diaminopimelic acid decarboxylase and arginyl-tRNA synthase.	<i>Corynebacterium glutamicum</i>	40,566 28-Jul-99	
rxa01014	2724	GB_BA1: MTV008	63033	AL021246	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 108/162.	<i>Mycobacterium tuberculosis</i>	56,167 17-Jun-98	
		GB_BA1: STMAMPEPN	2849	L23172	<i>Streptomyces lividans</i>	<i>Streptomyces</i>	57,067 18-MAY-1994	
		GB_BA1: SC7H2	42655	AL109732	<i>Streptomyces coelicolor</i> cosmid 7H2.	A3(2)	37,551 2-Aug-99	
rxa01022	1203	GB_PAT: A68384	1080	A68384	Sequence 1 from Patent WO9748809.	<i>Mycobacterium avium</i>	56,913 06-MAY-1999	
		GB_BA2: AF077728	1346	AF077728	<i>Mycobacterium smegmatis</i> D-alanine ligase gene, complete cds.	<i>Mycobacterium smegmatis</i>	57,203 1-Jan-99	

TABLE 4-continued
ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology	Date of Deposit (GAP)
rxa01055		GB_BA1: MSGB1723CS	38477	L78825	<i>Mycobacterium leprae</i> cosmid B1723 DNA sequence.	<i>Mycobacterium leprae</i>	54,599	15-Jun-96
rxa01056	1023	GB_BA2: AE001715	11086	AF001715	<i>Thermatoga maritima</i> section 27 of 136 of the complete genome.	<i>Thermatoga maritima</i>	39,034	2-Jun-99
		GB_EST38: AW046857	161	AW046857	UI-M_BH1_aka-a-04-0-U1\$1 NIH_BMAP_M_S2Mus musculus cDNA clone	<i>Mus musculus</i>	45,963	18-Sep-99
		GB_EST38: AW049435	244	AW049435	UI-M_BH1_aka-a-04-0-U1\$1 mRNA sequence.		40,984	18-Sep-99
rxa01057	1626	GB_PL1: LPAJ5046	656	AJ225046	UI-M_BH1_ams-b-01-0-U1\$1 NIH_BMAP_M_S2Mus musculus cDNA sequence.		37,117	22-Jul-98
		GB_PL2: SPAC806	22870	AL117212	<i>Lycopersicon peruvianum</i> mRNA for Hsp20.1 protein.	<i>Schizosaccharomyces pombe</i>	38,211	24-Nov-99
		GB_PL2: SPAC806	22870	AL117212	<i>S. pombe</i> chromosome I cosmid c806.	<i>Schizosaccharomyces pombe</i>	36,934	24-Nov-99
rxa01082	783	GB_BA2: AF112535	4363	AF112535	<i>Corynebacterium glutamicum</i> putative glutaredoxin NrdH (nrdH), NrdI (nrdI), and ribonucleotide reductase alpha-chain (nrdE) genes, complete cds.	<i>Corynebacterium glutamicum</i>	99,794	5-Aug-99
		GB_PL2: TAE237897	8020	A1237897	<i>Triticum aestivum</i> sbel gene, exons 1-14.	<i>Triticum aestivum</i>	37,132	1-Nov-99
		GB_PL2: AF076680	10499	AF076680	<i>Aegilops tauschii</i> starch branching enzyme-I (SBE-I) gene, complete cds.	<i>Aegilops tauschii</i>	38,651	14-MAY-1999
rxa01113	260	GB_VI: ASU02468	11424	U02468	African swine fever virus BA71V (A489R, A280R, A505R, A498R, A528R, A506R, and A542R) genes, complete cds.	African swine fever virus	31,923	28-Apr-94
		GB_VI: ASU18466	170101	U18466	African swine fever virus, complete genome.		31,923	22-Apr-95
		GB_GSS5: AQ752779	1647	AQ752779	HS_5569_B1_D02_SP6 RICl-11 Human Male BAC Library <i>Homo sapiens</i> genomic clone Plate = 1145 Col = 3 Row = H, genomic survey sequence.	<i>Homo sapiens</i>	37,154	19-Jul-99
rxa01115	876	GB_BA1: AB014757	6057	AB014757	<i>Pseudomonas</i> sp. 61-3 ketothiolase and PHB synthase, complete cds.	<i>Pseudomonas</i> sp. 61-3	40,850	26-DEC-1998
		GB_IN2: DMU60591	5630	U60591	<i>Drosophila melanogaster</i> (Kuzhanai) mRNA, complete cds.	<i>Drosophila melanogaster</i>	37,326	10-Sep-96
		GB_RO: MMMMP10	1744	Y13185	<i>Mus musculus</i> mRNA for stromelysin-2.	<i>Mus musculus</i>	35,877	14-Jan-98
		GB_BA1: SC4C6	30941	AL079355	<i>Streptomyces coelicolor</i> cosmid 4C6.	<i>Streptomyces coelicolor</i>	40,616	21-Jun-99
		GB_BA2: AF109386	6551	AF109386	<i>Streptomyces</i> sp. 2065 protocatechuaic acid catabolic gene cluster, complete sequence.	<i>Streptomyces</i> sp. 2065	64,099	06-DEC-1999
rxa01117	864	GB_BA1: MTCY07A7	23967	Z95556	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 109/162.	<i>Mycobacterium tuberculosis</i>	41,716	17-Jun-98
		GB_BA2: AF109386	6551	AF109386	<i>Streptomyces</i> sp. 2065 protocatechuaic acid catabolic gene cluster, complete sequence.	<i>Streptomyces</i> sp. 2065	62,116	06-DEC-1999
		GB_BA2: AF003947	5475	AF003947	<i>Rhodococcus opacus</i> succinyl CoA: 3-oxoadipate CoA transferase subunit homolog (pcatF) gene, partial cds, protocatechuate dioxygenase beta subunit (pcatF) gene, partial cds, and 3-oxoadipyl CoA thiolase homolog (pcatF') gene, partial cds.	<i>Rhodococcus opacus</i>	36,712	12-MAR-1998
					enol-lactone hydrolase/cytochrome c oxidase subunit (pcatC), 3-carboxy-3-hydroxy-3-muconolactone decarboxylase (pcatB), 3-oxoadipate PeaR (pcatR) genes, complete cds, and 3-oxoadipyl CoA thiolase homolog (pcatF') gene, partial cds.			
					<i>Xanthomonas campestris</i> genes and orfX.			
rxa01120	1401	GB_BA1: XCLPSU	2578	Y11313	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 108/162.	<i>Xanthomonas campestris</i>	39,833	20-Jan-98
		GB_BA1: MTY008	63033	AL021246	<i>Caulobacter crescentus</i> partial tig gene and ctp; ctaA, ctpX, ion genes.	<i>Mycobacterium tuberculosis</i>	36,715	17-Jun-98
		GB_BA1: CAJ10321	6710	AJ10321	<i>Azospirillum brasilense</i> trigger factor (tig), heat-shock protein ClpP (clpP),	<i>Caulobacter crescentus</i>	63,311	01-OCT-1998
		GB_BA2: AF150957	4440	AF150957	and heat-shock protein ClpX (clpX) genes, complete cds; and Lon protease (lon) gene, partial cds.	<i>Azospirillum brasilense</i>	60,613	7-Jun-99

TABLE 4-continued
ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit		% homology (GAP)	Date of Deposit
rxa01126	583	GB-HTG3: AC009199	66498	AC009199	<i>Drosophila melanogaster</i> chromosome 2 clone BACR10J23 (D1024) RPCI-98 10J.23 map 37B-37B strain Y; cn bw sp; *** SEQUENCING IN PROGRESS ***; 79 unordered pieces.	<i>Drosophila melanogaster</i>	35,294	20-Sep-99
		GB-HTG3: AC009199	66498	AC009199	<i>Drosophila melanogaster</i> chromosome 2 clone BACR10J23 (D1024) RPCI-98 10J.23 map 37B-37B strain Y; cn bw sp; *** SEQUENCING IN PROGRESS ***; 79 unordered pieces.	<i>Drosophila melanogaster</i>	35,294	20-Sep-99
GB_PL1: AB016380			81284	AB016380	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone; MTG10, complete sequence.	Arabidopsis thaliana	34,477	20-Nov-99
GB_BAI: MLCB22	980	GB_BAI: MTCY190	40281	298741	<i>Mycobacterium leprae</i> cosmid B22.	<i>Mycobacterium leprae</i>	61,570	22-Aug-97
GB_BAI: SC5F7			34150	270283	<i>Mycobacterium tuberculosis</i> H37RV complete genome; segment 98/162.	<i>Mycobacterium tuberculosis</i>	60,434	17-Jun-98
			40024	AL096872	<i>Sireptomyces coelicolor</i> cosmid 5F7.	<i>Sireptomyces coelicolor</i>	57,011	22-Jul-99
rxa01181		GB_EST3: H01832	381	H01832	yJ28c11.s1 Soates placenta Nb2HP <i>Homo sapiens</i> cDNA clone IMAGE: 150068 3', mRNA sequence.	<i>Homo sapiens</i>	41,406	19-Jun-95
GB_PRA: AC004850		GB_GSS11: AQ304150	105891	AC004850	<i>Homo sapiens</i> PAC clone D10665C04 from 7p14-p13, complete sequence.	<i>Homo sapiens</i>	37,428	26-Feb-99
			528	AQ304150	HS_3208_A1_D12_T7 CIT Approved Human Genomic Sperm Library D	<i>Homo sapiens</i>	37,421	16-DEC-1998
rxa01236					<i>Homo sapiens</i> genomic clone Plate = 32 Row = G, genomic survey sequence.			
GB_BAI: MIV025	1392	GB_BAI: MSGB577COS	121125	AL022121	<i>Mycobacterium tuberculosis</i> H37RV complete genome; segment 155/162.	<i>Mycobacterium tuberculosis</i>	58,315	24-Jun-99
GB_BAI: MLCB2407		GB_BAI: X91182	37770	L01263	<i>M. leprae</i> genomic dna sequence, cosmid b577.	<i>Mycobacterium leprae</i>	56,323	14-Jun-96
GB_BAI: SPIN112D	1278	GB_BAI: SPIN112D	35615	AL023596	<i>Mycobacterium leprae</i> cosmid B2407.	<i>Mycobacterium leprae</i>	37,645	27-Aug-99
GB_BAI: EST7: W93397			345	X91182	Bacterial sp. partial 16S rRNA gene (clone group G10).	unidentified bacterium	41,228	15-Jul-96
			347	269277	Bacterial sp. partial 16S rRNA gene (clone group IN12d).	Bacteria	38,905	24-Jun-98
			545	W93397	zd95b03.s1 Soates_fetal_heart_NbHH19W <i>Homo sapiens</i> cDNA clone IMAGE: 357197 3', mRNA sequence.	<i>Homo sapiens</i>	40,516	25-Nov-96
rxa01277	2127	GB_PL2: AF111709	52684	AF111709	<i>Oryza sativa</i> subsp <i>indica</i>		37,410	26-Apr-99
GB_JN1: CELZC250		GB_EST1: Z14808	34372	AF003383	Retrosat 2 retrotransposon, complete sequences; and unknown genes.		35,506	14-MAY-1997
GB_VI: S62819	498		331	Z14808	<i>Caenorhabditis elegans</i> cDNA library <i>Caenorhabditis elegans</i> clone cm5e4 5', mRNA sequence.		36,890	19-Jun-97
			3348	S62819	F2L = putative RNA polymerase-associated transcription factor ... F4R = type I orf virus topoisoerase homolog [orf virus OV, N22, host = sheep, Genomic, 3 genes, 3348 nt].		40,471	25-Aug-93
GB_PRA: HUMCCLECI			17079	AF077344	<i>Homo sapiens</i> cartilage-derived C-type lectin (CLECSF1) gene, exons 1 and 2.	<i>Homo sapiens</i>	34,631	15-OCT-1999
GB_PRA: HUMCCLECI			17079	AF077344	<i>Homo sapiens</i> cartilage-derived C-type lectin (CLECSF1) gene, exons 1 and 2.	<i>Homo sapiens</i>	39,300	15-OCT-1999
rxa01354	1059	GB_PRI: D87675	301692	D87675	<i>Homo sapiens</i> DNA for amyloid precursor protein, complete cds.	<i>Homo sapiens</i>	37,984	22-Sep-97
GB_PRI: D87675		GB_RO: MMNUCLEO	301692	D87675	<i>Homo sapiens</i> DNA for amyloid precursor protein, complete cds.	<i>Homo sapiens</i>	35,140	22-Sep-97
GB_BAI: MTCY71		11478	11478	X07699	Mouse nucleolin gene.	<i>Mus musculus</i>	37,146	27-Aug-98
GB_BAI: ACCPSXM		42729	42729	Z92271	<i>Mycobacterium tuberculosis</i> H37RV complete genome; segment 141/162.	<i>Mycobacterium tuberculosis</i>	39,496	10-Feb-99
GB_BA2: ECU05248	1781	2748	1781	X81320	<i>A. calcoaceticus</i> epsX and epsM genes.	<i>Achromobacter calcoaceticus</i>	40,333	19-OCT-1994
			1781	U05248	<i>Escherichia coli</i> polysaccharide acid gene cluster region 2 (neuD and neuB) genes, complete cds.	<i>Escherichia coli</i>	34,995	1-Feb-95

TABLE 4-continued
ALIGNMENT RESULTS

ID #	Length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa01385	2004	GB_BA1: FVBPENTA	2519	M98557	<i>Flavobacterium</i> sp. pentachlorophenol 4-monoxygenase gene, complete mRNA.	<i>Flavobacterium</i> sp.	40,855 26-Apr-93	
		GB_PAT: I19994	2516	119994	Sequence 2 from patent US 5512478.	Unknown.	40,855 07-OCT-1996	
		GB_BA2: AF059680	2410	AF059680	<i>Sphingomonas</i> sp. UG30 pentachlorophenol 4-monoxygenase (pcpB) gene, complete cds; and pentachlorophenol 4-monoxygenase reductase (pcpD) gene, partial cds.	<i>Sphingomonas</i> sp. UG30	42,993 27-Apr-99	
rxa01426	750	GB_GSS3: B35912	313	B35912	<i>Homo sapiens</i> genomic clone Plate = CT 811 Col = 4 Row = G, genomic survey sequence.	<i>Homo sapiens</i>	38,019 17-OCT-1997	
		GB_GSS1: FR0027767	497	AL020589	<i>F. rubripes</i> GS5 sequence, clone 197B17a3, genomic survey sequence.	<i>Fugu rubripes</i>	35,814 10-DEC-1997	
		GB_GSS5: AQ774340	449	AQ774340	HS_3137_A2_E11_MR CTT Approved Human Genomic Sperm Library D	<i>Homo sapiens</i>	40,535 29-Jul-99	
rxa01427	1044	GB_BA2: AF036766	3487	AF036766	<i>Lactobacillus reuteri</i> plasmid pTE15 replication-associated protein A (repA) and replication-associated protein B (repB) genes, complete cds.	<i>Lactobacillus reuteri</i>	39,101 19-Feb-98	
		GB_PR4: AC007032	124803	AC007032	<i>Homo sapiens</i> clone NH00228N19, complete sequence.	<i>Homo sapiens</i>	34,180 17-Jul-99	
		GB_PR4: AC007032	126803	AC007032	<i>Homo sapiens</i> clone NH00228N19, complete sequence.	<i>Homo sapiens</i>	36,858 17-Jul-99	
rxa01428	1260	GB_BA1: SCH24	41625	AL049826	Streptomyces coelicolor thioredoxin H24.	<i>Streptomyces coelicolor</i>	51,278 11-MAY-1999	
		GB_BA2: AF031590	6676	AF031590	Streptomyces coelicolor thioredoxin (trxA) gene, partial cds; SpoQJ-like, GidB-like, Jag-like, inner membrane protein, and 9-10 kDa protein-like genes, complete cds; RNase P protein (rnpA) gene, partial cds; and unknown gene.	<i>Streptomyces coelicolor</i>	39,389 20-Feb-98	
		GB_BA1: SCTRXXARNP	6676	Y16311	<i>Streptomyces coelicolor</i> trNA & rnpA genes & ORFs 205, 244, 255, 239, v139b08_y1_124.	<i>Streptomyces coelicolor</i>	39,389 18-DEC-1998	
rxa01430	1311	GB_EST30: AL643302	254	AI643302	v139b08_y1_124. Stratogene mouse skin (#937313) <i>Mus musculus</i> cDNA clone IMAGE: 974533 5' similar to SW: 6PGD_HUMAN P52209_6- PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING, mRNA sequence.	<i>Mus musculus</i>	38,627 29-Apr-99	
		GB_EST34: AI788121	490	AI788121	IMAGE: 2087835 5' similar to SW: 6PGD_HUMAN P52209_6- PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING, mRNA sequence.	<i>Mus musculus</i>	40,583 2-Jul-99	
		GB_EST16: AA560354	253	AA560354	v139b08_r1 Stratogene mouse skin (#937313) <i>Mus musculus</i> cDNA clone IMAGE: 974533 5' similar to TR: G084325 G984325.	<i>Mus musculus</i>	42,544 18-Aug-97	
		GB_EST22: AI069195	892	AI069195	PHOSPHOGLUCONATE DEHYDROGENASE,, mRNA sequence.		40,964 09-DEC-1999	
		GB_EST26: AI392390	574	AI392390	mage0005df02/Magnaporthe grisea Apressorium Stage cDNA Library	<i>Pyricularia grisea</i>	40,127 3-Feb-99	
		GB_HTG2: AC004845	140230	AC004845	NCSCL1B12T Subtracted CandidaNeurospora crassa cDNA clone SC1B12 3' similar to adenylyl kinase 2 (ATP-AMP transphosphorylase), mRNA sequence.	<i>Neurospora crassa</i>		
					<i>Homo sapiens</i> clone D10635C05, *** SEQUENCING IN PROGRESS ***, 7 <i>Homo sapiens</i> unordered pieces.		36,437 12-Jun-98	

TABLE 4-continued
ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa01437	1506	GB_BA1: CGPTAAACKA GB_BA1: MTCY22610 GB_HTG3: AC010254	3657 / X89084 35420 / Z84724 114363 / AC010254	X89084 Z84724 AC010254	<i>C. glutamicum</i> pia gene and ackA gene. <i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 21/162. <i>Homo sapiens</i> chromosome 5 clone C1-HSPC_434Q11, *** SEQUENCING IN PROGRESS **, 58 unordered pieces.	<i>Corynebacterium glutamicum</i> <i>Mycobacterium tuberculosis</i> <i>Homo sapiens</i>	100,000 23-MAR-1999 54,867 17-Jun-98 35,547 15-Sep-99	
rxa01461	735	GB_BA2: AF003947	5475	AF003947	Rhodococcus opacus CoA: 3-oxoadipate CoA transferase subunit homolog (pcal) gene, partial cds, protocatechuate dioxygenase beta subunit (pcalH), protocatechuate dioxygenase alpha subunit (pcalG), 3-carboxy-cis, cis-muconate cycloisomerase homolog (pcalB), 3-oxoadipate PcaR (pcarR) genes, complete cds, and 3-oxoadipyl CoA thiolase homolog (pcalF) gene, partial cds.	Rhodococcus opacus	57,939 12-MAR-1998	
			187408	AL078638	Human DNA sequence from clone RP11-535K18 on chromosome X(26,2-27,1, complete sequence. X(26,2-27,1, complete sequence. Uf-R-Y0-abw-e-02-0-Ul,s2 Uf-R-Y0Rattus norvegicus cDNA clone Uf-R-Y0- abw-e-02-0-Ul , mRNA sequence.	<i>Homo sapiens</i>	37,123 22-Nov-99	
		GB_EST33: AI764654	420	A1764654	<i>Rhodococcus opacus</i> succinyl CoA: 3-oxoadipate CoA transferase subunit homolog (pcal) gene, partial cds, protocatechuate dioxygenase beta subunit (pcalH), protocatechuate dioxygenase alpha subunit (pcalG), 3-carboxy-cis, cis-muconate cycloisomerase homolog (pcalB), 3-oxoadipate PcaR (pcarR) genes, complete cds, and 3-oxoadipyl CoA thiolase homolog (pcalF) gene, partial cds.	<i>Rattus norvegicus</i>	35,885 25-Jun-99	
rxa01462	813	GB_BA2: AF003947	5475	AF003947	Rhodococcus opacus CoA: 3-oxoadipate CoA transferase subunit homolog (pcal) gene, partial cds, protocatechuate dioxygenase beta subunit (pcalH), protocatechuate dioxygenase alpha subunit (pcalG), 3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylase (pcalB) and PcaR (pcarR) genes, complete cds, and 3-oxoadipyl CoA thiolase homolog (pcalF) gene, partial cds.	Rhodococcus opacus	66,667 12-MAR-1998	
		GB_BA1: SC4C6 GB_BA2: AF109386	30941	AL079355 6551 AF109386	<i>Streptomyces coelicolor</i> cosmid 4C6. <i>Streptomyces</i> sp. 2065 protocatechuaic acid catabolic gene cluster, complete sequence.	<i>Streptomyces</i> sp. 2065	40,822 21-Jun-99 56,049 06-DEC-1999	
rxa01464	414	GB_BA1: AB009343	6342	AB009343	<i>Fratervia</i> sp. ANA-18 ORFR2, catB1, catC1, catD1 genes, complete cds.	<i>Fratervia</i> sp. ANA-18	50,966 26-MAY-1999	
		GB_GSS10: AQ241375	284	AQ241375	CITB1-EI-250507.TF1 CITB1-EI <i>Homo sapiens</i> genomic clone 250507T, genomic survey sequence.	<i>Homo sapiens</i>	39,085 30-Sep-98	
		GB_HTG3: AC010363	174962	AC010363	<i>Homo sapiens</i> chromosome 5 clone CITB-H1_2039P12, *** SEQUENCING IN PROGRESS **, 43 unordered pieces.	<i>Homo sapiens</i>	35,784 15-Sep-99	
rxa01465	1284	GB_BA1: ROX99622 GB_BA1: D83237 GB_EST9: AA119571	7224 X99622 1626 D83237 4445 AA119571	X99622 D83237 AA119571	<i>Rhodococcus opacus</i> catR, catA, catB, catC genes and five ORFs. <i>Rhodococcus erythropolis</i> DNA for catechol 1,2-dioxygenase, complete cds. mp0604_11 Soares 2NbMT <i>Mus musculus</i> cDNA clone IMAGE: 574375 5' similar to TR: G559375 G559375 RAS GTPASE-ACTIVATING PROTEIN;, mRNA sequence. wg17906_x1 Soares_NFL_T_GBC_S1 <i>Homo sapiens</i> cDNA clone IMAGE: 228371 3', mRNA sequence.	<i>Rhodococcus erythropolis</i> <i>Mus musculus</i>	58,814 24-Sep-97 53,904 1-Sep-99 39,551 17-Feb-97	
rxa01466	1083	GB_EST37: AI934978	425	AI934978	aa32g06_s1 NFL CGAP_GCB1 <i>Homo sapiens</i> cDNA clone IMAGE: 815002, aa32g06_s1 NFL CGAP_GCB1 <i>Homo sapiens</i> cDNA clone IMAGE: 1759280 3', mRNA sequence.	<i>Homo sapiens</i>	43,609 2-Sep-99	
		GB_EST15: AA4465729	289	AA4465729	aa32g06_s1 NFL CGAP_GCB1 <i>Homo sapiens</i> cDNA clone IMAGE: 815002, similar to TR: Q99988 Q99988 TGF-BETA cDNA clone IMAGE: 1759280 3', similar to TR: Q99988 Q99988 TGF-BETA SUPERFAMILY PROTEIN. [1], mRNA sequence.	<i>Homo sapiens</i>	41,115 13-Aug-97	
		GB_EST24: AI219091	633	AI219091	cg12a05_x1 Soares_placenta_8109weeks_21NHP8109 W <i>Homo sapiens</i> cDNA clone IMAGE: 1759280 3', similar to TR: Q99988 Q99988 TGF-BETA SUPERFAMILY PROTEIN. [1], mRNA sequence.	<i>Homo sapiens</i>	36,066 29-Nov-98	

TABLE 4-continued

ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa01477	1671	GB_BA2: CGU89648	1105	U89648	<i>Corynebacterium glutamicum</i> unidentified sequence involved in histidine biosynthesis, partial sequence vx11g06r1 Soares 2NBMT <i>Mus musculus</i> cDNA clone IMAGE: 1264186 5' similar to gb: M73696 Murine Givr-1 mRNA, complete cds (MOUSE), mRNA sequence.	<i>Corynebacterium glutamicum</i>	49,776 30-MAR-1999	
		GB_EST21: AA919685	782	AA919685	<i>Homo sapiens</i> chromosome 20 clone RP5-1005F21, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Mus musculus</i>	37,762 20-Apr-98	
		GB-HTG2: HS1005F21	101795	AL078633	<i>Homo sapiens</i> chromosome 20 clone RP5-1005F21, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	38,371 30-Nov-99	
rxa01499	3945	GB_PRA4: AC006454	153201	AC006454	<i>Homo sapiens</i> clone D10852P06, complete sequence.	<i>Homo sapiens</i>	38,033 13-Aug-99	
		GB_BA1: LSIXSYNT	4724	AC006454	<i>Lysobacter</i> sp. gene encoding synthetase.	<i>Lysobacter</i>	42,840 8-Jan-97	
		GB_PRA4: AC006454	153201	AC006454	<i>Homo sapiens</i> clone D10852P06, complete sequence.	<i>Homo sapiens</i>	38,823 13-Aug-99	
rxa01502	1356	GB_PAT: I92046	2203	I92046	Sequence 13 from patient US 57262/99.	Unknown.	39,755 01-DEC-1998	
		GB_PAT: I78757	2203	I78757	Sequence 13 from patient US 5693781.	Unknown.	39,755 3-Apr-98	
		GB_BA1: MTCY359	283859		<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 84/162.	<i>Mycobacterium tuberculosis</i>	36,613 17-Jun-98	
		GB_BA1: SCE9	36021	A049841	<i>Streptomyces coelicolor</i> cosmid E9.	<i>Streptomyces coelicolor</i>	60,637 19-MAY-1999	
rxa01509	597	GB_BA1: MTY15C10	37730	A1049841	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 154/162.	<i>Mycobacterium tuberculosis</i>	59,296 17-Jun-98	
		GB_BA1: MLCB2548	33050	Z95436	<i>Mycobacterium leprae</i> cosmid B2548.	<i>Mycobacterium leprae</i>	59,764 27-Aug-99	
rxa01510	1404	GB_GSS9: AQ129927	38916	A1023093	HS_2165_B1_D09 MR CIT Approved Human Genomic Sperm Library D	<i>Homo sapiens</i>	36,136 23-Sep-98	
			440	AQ129927	<i>Homo sapiens</i> genomic clone Plate = 2165 Col = 17 Row = H, genomic survey sequence.			
		GB_BA2: AF016585	41097	AF016585	<i>Streptomyces caelstis</i> cytochrome P-450 hydroxylase homolog (nidi) gene, partial cds; polyketide synthase modules 1 through 7 (nida) genes, complete cds, and N-methyltransferase homolog gene, partial cds.	<i>Streptomyces caelstis</i>	37,464 07-DEC-1997	
		GB-HTG4: AC010747	216500	AC010747	<i>Homo sapiens</i> chromosome unknown clone NH0555H109, WORKING DRAFT SEQUENCE, in unordered pieces.	<i>Homo sapiens</i>	33,022 29-OCT-1999	
rxa01511	1065	GB_BA1: BRLBIOBA	1647	D14084	<i>Brevibacterium flavum</i> gene for biotin synthetase, complete cds.	<i>Corynebacterium glutamicum</i>	40,283 3-Feb-99	
		GB_GSS3: B45213	358	B45213	HS_1060_B2-D07-MF_abi CIT Human Genomic Sperm Library C <i>Homo sapiens</i> genomic clone Plate = CT 782 Col = 14 Row = H, genomic survey sequence.	<i>Homo sapiens</i>	49,505 21-OCT-1997	
		GB-HTG4: AC010747	216500	AC010747	<i>Homo sapiens</i> chromosome unknown clone NH0555H109, WORKING DRAFT SEQUENCE, in unordered pieces.	<i>Homo sapiens</i>	33,819 29-OCT-1999	
rxa01513	2682	GB_BA1: MTCY7H7B	24244	Z95557	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 153/162.	<i>Mycobacterium smegmatis</i>	40,354 18-Jun-98	
		GB_BA2: AF037269	2364	AF037269	<i>Mycobacterium smegmatis</i> cell division protein (FisH) gene, complete cds.	<i>Mycobacterium leprae</i>	60,814 19-Aug-98	
		GB_BA1: MLCB2548	38916	A1023093	<i>Mycobacterium leprae</i> cosmid B2548.	<i>Mycobacterium leprae</i>	39,992 27-Aug-99	
rxa01593	990	GB_BA1: U00012	33312	U00012	<i>Mycobacterium leprae</i> cosmid B1308.	<i>Caenorhabditis elegans</i>	39,126 30-Jan-96	
		GB_INI: CELF27E11	25700	AF016413	<i>Caenorhabditis elegans</i> cosmid F27E11.	<i>Caenorhabditis elegans</i>	34,227 2-Aug-97	
		GB_OV: DYGAAGR	4354	L01423	<i>Discosyne ornata</i> (clone OL4) agrin mRNA, 3' end cds.	<i>Discosyne ornata</i>	38,414 28-Apr-93	
rxa01608	1962	GB_BA2: AF119150	18605	AF119150	<i>Vibrio cholerae</i> Rtx toxin gene cluster, complete cds.	<i>Vibrio cholerae</i>	36,919 21-MAR-1999	
		GB_BA2: AF119150	18605	AF119150	<i>Vibrio cholerae</i> Rtx toxin gene cluster, complete cds.	<i>Vibrio cholerae</i>	38,130 21-MAR-1999	
rxa01620	3441	GB_PR3: HSS2D1	148691	Z96811	Human DNA sequence from PAC 52D1 on chromosome Xq21. Contains CA repeats, STS.	<i>Homo sapiens</i>	35,501 23-Nov-99	
rxa01640		GB_BA2: AF079155	686	AF079155	Ralstonia eutropha phasin (phap) mRNA, complete cds.	<i>Ralstonia eutropha</i>	40,497 6-Apr-99	
		GB_IN2: AF039570	1866	AF039570	<i>Caenorhabditis elegans</i> aryl hydrocarbon receptor ortholog AHR-1 (ahr-1) mRNA, complete cds.	<i>Caenorhabditis elegans</i>	39,699 04-OCT-1999	

TABLE 4-continued
ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homolog (GAP)	Date of Deposit
rxa01653	1584	GB-HTG7: AC010997	187768	AC010997	<i>Homo sapiens</i> clone RP11-399K21, *** SEQUENCING IN PROGRESS ***, <i>Homo sapiens</i>		34,516	08-DEC-1999
		GB-HTG7: AC010997	187768	AC010997	<i>Homo sapiens</i> clone RP11-399K21, *** SEQUENCING IN PROGRESS ***, <i>Homo sapiens</i>		36,177	08-DEC-1999
		GB_VI: AF030154	34446	AF030154	Bovine adenovirus 3 complete genome.	bovine adenovirus type 3	40,345	27-Jan-99
		GB_BAI: AB010645	16836	AB010645	<i>Acetobacter xylinus</i> genes for endoglucanase, cellulose synthase subunit ABCD and beta-glucosidase, complete cds.	<i>Acetobacter xylinus</i>	34,783	13-Feb-99
		GB_BAI: AB010645	16836	AB010645	<i>Acetobacter xylinus</i> genes for endoglucanase, cellulose synthase subunit ABCD and beta-glucosidase, complete cds.	<i>Acetobacter xylinus</i>	37,598	13-Feb-99
rxa01716	509							
rxa01728	1098	GB_BA1: ABCBCSABCD	9540	M37202	<i>A. xylinum</i> bes A, B, C and D genes, complete cds'.	<i>Acetobacter xylinus</i>	39,173	24-Apr-93
		GB_BA2: CORCSLYS	2821	M89931	<i>Corynebacterium glutamicum</i> beta C-S lyase (aecD) and branched-chain amino acid uptake carrier (fmoQ) genes, complete cds, and hypothetical protein Yhbw (yhbw) gene, partial cds.	<i>Corynebacterium glutamicum</i>	99,636	4-Jun-98
		GB_PL2: HAAP	931	X95952	<i>H. annuus</i> mRNA for aquaporin.	<i>Helianthus annuus</i>	39,231	14-Jul-99
		GB-HTG1: CEY32F6	187816	AL008875	<i>Caenorhabditis elegans</i> chromosome V clone Y32F6, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Caenorhabditis elegans</i>	37,431	9-Nov-97
rxa01732	1173	GB_PR4: HUAC004125	194020	AC004125	<i>Homosapiens</i> Chromosome 16 BAC clone CIT987SK-625P11, complete sequence.	<i>Homosapiens</i>	35,345	23-Nov-99
		GB_PR4: HUAC004125	194020	AC004125	<i>Homosapiens</i> Chromosome 16 BAC clone CIT987SK-625P11, complete sequence.	<i>Homosapiens</i>	37,381	23-Nov-99
rxa01810	1200	GB_IN1: CER11A5	26671	Z83122	<i>Caenorhabditis elegans</i> cosmid R11A5, complete sequence.	<i>Caenorhabditis elegans</i>	36,140	2-Sep-99
		GB_EST28: AL499508	403	AI499508	to02d01.x1 NCL CGAP_U12 <i>Homosapiens</i> cDNA clone IMAGE: 2177857 3' similar to SW. NU4M_PANTRE P03906 NU4M_UBIQUINONE OXIDOREDUCTASE CHAIN 4 _b , mRNA sequence.	<i>Homosapiens</i>	36,725	11-MAR-1999
		GB_EST28: AL499508	403	AI499508	to02d01.x1 NCL CGAP_U12 <i>Homosapiens</i> cDNA clone IMAGE: 2177857 3' <i>Homosapiens</i> OXIDOREDUCTASE CHAIN 4 _b , mRNA sequence.	<i>Homosapiens</i>	38,264	11-MAR-1999
rxa01828	1545	GB_BA1: MLCB1770	37821	Z70722	<i>Mycobacterium leprae</i> cosmid B1770.	<i>Mycobacterium leprae</i>	36,411	29-Aug-97
		GB-HTG2: AC008073	173144	AC008073	<i>Homosapiens</i> clone NH0507M03, *** SEQUENCING IN PROGRESS ***, 3 <i>Homosapiens</i>	<i>Homosapiens</i>	36,310	17-Jul-99
		GB-HTG2: AC008073	173144	AC008073	<i>Homosapiens</i> clone NH0507M03, *** SEQUENCING IN PROGRESS ***, 3 <i>Homosapiens</i>	<i>Homosapiens</i>	36,310	17-Jul-99
rxa01829	1446	GB_IN1: AB018544	620	AB018544	<i>Hydra magnipapillata</i> mRNA for Hym-176 preprohormone, complete cds.	<i>Hydra magnipapillata</i>	34,855	6-Feb-99
		GB_EST8: AA003136	450	AA003136	mg-1e01.rl Soares mouse embryo NbME13.5 14.5 <i>Mus musculus</i> cDNA clone [HUMAN], mRNA sequence.	<i>Mus musculus</i>	42,202	19-Jul-96
		GB_IN1: AB018544	620	AB018544	<i>Hydra magnipapillata</i> mRNA for Hym-176 preprohormone, complete cds.	<i>Hydra magnipapillata</i>	35,968	6-Feb-99
		GB_BAI: MTV033	21620	AL021928	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 11/162.	<i>Mycobacterium tuberculosis</i>	38,679	17-Jun-98
		GB_BAI: MLCL622	42498	Z95398	<i>Mycobacterium leprae</i> cosmid 1622.	<i>Mycobacterium leprae</i>	38,911	24-Jun-97
		GB_BAI: MSCB983CS	36788	L78828	<i>Mycobacterium leprae</i> cosmid B983 DNA sequence.	<i>Mycobacterium leprae</i>	38,933	15-Jun-96
rxa01934	681	GB_PR4: DJ534K4	216387	AF09907	<i>Homosapiens</i> S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds.	<i>Homosapiens</i>	39,189	23-DEC-1998
		GB-HTG2: AC006342	201618	AC006342	<i>Homosapiens</i> clone DJ054D12, *** SEQUENCING IN PROGRESS ***, 3 <i>Homosapiens</i>	<i>Homosapiens</i>	34,412	11-Jan-99
					unordered pieces.			

TABLE 4-continued

ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
		GB-HTG2: AC006342	201618 AC006342	<i>Homo sapiens</i> clone DJ0054D12, *** SEQUENCING IN PROGRESS ***, 3 unorderd pieces.	<i>Homo sapiens</i>	34,412 11-Jan-99	
rxa01967	1266	GB_IN2: AC005467	62091 AC005467	<i>Drosophila melanogaster</i> , chromosome 2R, region 48C1-48C2, P1 clone DS00568, complete sequence.	<i>Drosophila melanogaster</i>	35,252 12-DEC-1998	
		GB_BA2: AE001678	13485 AE001678	<i>Chlamydia pneumoniae</i> section 94 of 103 of the complete genome.	<i>Chlamydophila pneumoniae</i>	35,203 08-MAR-1999	
		GB_IN2: AC005467	62091 AC005467	<i>Drosophila melanogaster</i> , chromosome 2R, region 48C1-48C2, P1 clone DS00568, complete sequence.	<i>Drosophila melanogaster</i>	34,699 12-DEC-1998	
rxa01993	1166	GB_BA1: PPVANAB GB-HTG2: AC006799	2864 Y14759 278007 AC006799	<i>Pseudomonas putida</i> vanA and vanB genes. <i>Caenorhabditis elegans</i> clone Y51H7, *** SEQUENCING IN PROGRESS ***, 7 unorderd pieces.	<i>Caenorhabditis elegans</i>	51,697 09-MAY-1998 38,455 23-Feb-99	
		GB-HTG2: AC006799	278007 AC006799	<i>Caenorhabditis elegans</i> clone Y51H7, *** SEQUENCING IN PROGRESS ***, 7 unorderd pieces.	<i>Caenorhabditis elegans</i>	38,455 23-Feb-99	
rxa01994	1098	GB-HTG4: AC009961	231522 AC009961	<i>Homo sapiens</i> chromosome unknown clone NH0357L02, WORKING DRAFT SEQUENCE, in unorderd pieces.	<i>Homo sapiens</i>	35,576 29-OCT-1999	
		GB-HTG4: AC009961	231522 AC009961	<i>Homo sapiens</i> chromosome unknown clone NH0357L02, WORKING DRAFT SEQUENCE, in unorderd pieces.	<i>Homo sapiens</i>	35,576 29-OCT-1999	
		GB-HTG4: AC009961	231522 AC009961	<i>Homo sapiens</i> chromosome unknown clone NH0357L02, WORKING DRAFT SEQUENCE, in unorderd pieces.	<i>Homo sapiens</i>	35,472 29-OCT-1999	
rxa01997	609	GB_BA1: SCH66 GB_EST29: AI558691	1798 AF112536 9153 AL049731 598 AI558691	<i>Corynebacterium glutamicum</i> ribonucleotide reductase beta-chain (ndf) gene, complete cds. <i>Streptomyces coelicolor</i> cosmid H66. fb79c10.y1 Zebrafish WASHU MPMIG EST Danio rerio cDNA 5' similar to SW-ATF3_HUMAN P1.8847 CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-3, mRNA sequence.	<i>Corynebacterium glutamicum</i> <i>Streptomyces coelicolor</i> <i>Danio rerio</i>	37,719 5-Aug-99 38,655 29-Apr-99 40,232 24-MAR-1999	
rxa02052	915	GB_EST3: R64206	453 R64206	y1.8d08.r1 Sora's placenta N22HP <i>Homo sapiens</i> cDNA clone IMAGE: 129595, mRNA sequence.	<i>Homo sapiens</i>	35,920 26-MAY-1995	
		GB_PR2: AC002540 GB_GSS3: B55001	70851 AC002540 406 B55001	Human BAC clone GS025M02 from 7q21-q22, complete sequence. C1THSP-385H2. TRB C1T-HSP <i>Homo sapiens</i> genomic clone 385H2.	<i>Homo sapiens</i>	37,099 12-Sep-97 35,599 20-Jun-98	
rxa02064	762	GB_PR4: AF135187 GB_PR3: AC005612	33016 AF135187 60904 AC005612	<i>Homo sapiens</i> interferon-induced protein p78 (MX1) gene, complete cds. <i>Homo sapiens</i> chromosome 21, P1 clone LB4# (LBNL HB), complete genomic survey sequence.	<i>Homo sapiens</i>	32,935 8-Jul-99 32,935 4-Sep-98	
		GB_PR1: HUM8DC11Z GB_BA1: MSGB32CS GB_BA1: MTCY338 GB_GSS10: AQ242118	3949 L35666 36404 L78818 29372 Z74697 766 AQ242118	<i>Homo sapiens</i> (subclone H8 10, f11 from P1 35 H5 C8) DNA sequence. <i>Mycobacterium leprae</i> cosmid B32 DNA sequence. <i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 127/162. 3123-4r <i>Ochrobactrum anthropi</i> BAC Library <i>Ochrobactrum anthropi</i> genomic clone 3123-4r, genomic survey sequence. <i>Homo sapiens</i> 12q22-103.4-106.5 BAC RPCII-718L23 (Roswell Park Cancer Institute Human BAC Library), complete sequence.	<i>Mycobacterium leprae</i> <i>Mycobacterium tuberculosis</i> <i>Ochrobactrum anthropi</i>	31,995 22-Aug-94 50,64 15-Jun-96 38,113 17-Jun-98 41,876 02-OCT-1998	
rxa02083	1533	GB_PR4: AC008055	196899 AC008055	<i>Homo sapiens</i> (subclone H8 10, f11 from P1 35 H5 C8) DNA sequence. Genomic sequence of Arabidopsis BAC RSA5, complete sequence. <i>Homo sapiens</i> 12q22-103.4-106.5 BAC RPCII-718L23 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	<i>Homo sapiens</i>	36,818 09-OCT-1999	
		GB_PL2: AC002292 GB_PR4: AC008055	120787 AC002292 196899 AC008055	Genomic sequence of Arabidopsis BAC RSA5, complete sequence. <i>Homo sapiens</i> 12q22-103.4-106.5 BAC RPCII-718L23 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	<i>Arabidopsis thaliana</i> <i>Homo sapiens</i>	37,517 02-OCT-1997 35,563 09-OCT-1999	
rxa02092	1761	GB_BA2: AF031929	2675 AF031929	<i>Lactobacillus helveticus</i> complete cds, and DNA mismatch repair enzyme (hexA) gene, partial cds.	<i>Lactobacillus helveticus</i>	36,149 8-Aug-98	

TABLE 4-continued
ALIGNMENT RESULTS

ID #	Length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
		GB-HTG1: HS0134F7	129811	AL049547	<i>Homo sapiens</i> chromosome 6 clone RP1-34F7, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	37,587	23-Nov-99
rxa02098	1869	GB_PR2: HSU24578 GB_BAI: CAJ10319	17488 5368	U24578 AJ010319	Human RP1 and complement C4B precursor (C4B) genes, partial cds. <i>Corynebacterium glutamicum</i> amnP, glnB, glnD genes and partial ftsY and fsp genes.	<i>Homo sapiens</i> <i>Corynebacterium glutamicum</i>	36,755 99,766	16-MAY-1996 14-MAY-1999
rxa02105	391	GB_EST17: GB_GSS6: AQ839377	352	AQ839377	EST00115 watermelon lambda zap express library <i>Citrullus lanatus</i> cDNA clone WMLS23.5 similar to translation initiation factor, mRNA sequence. HS_4640_B2_F09_T7A CTF Approved Human Genomic Sperm Library D	<i>Citrullus lanatus</i> <i>Homo sapiens</i>	37,231 37,500	10-Nov-97 30-Aug-99
rxa02111	1407	GB_BAI: SC6G10 GB_BAI: U00010 GB_BAI: MTICY336 GB-HTG2: AC007164	36734 41171 3295386 158320	AL049497 U00010 MTICY336 AC007164	<i>Streptomyces coelicolor</i> cosmid 6G10. <i>Mycobacterium leprae</i> cosmid B1170. <i>Mycobacterium tuberculosis</i> H37Rv complete genome: segment 70/162. <i>Homo sapiens</i> clone NH0304A10, *** SEQUENCING IN PROGRESS *** ; 3 unorderd pieces.	<i>Streptomyces coelicolor</i> <i>Mycobacterium leprae</i> <i>Mycobacterium tuberculosis</i> <i>Homo sapiens</i>	50,791 37,563 39,504 38,377	24-MAR-1999 01-MAR-1994 24-Jun-99 23-Apr-99
rxa02118	465	GB_PL2: PUMCDC2A GB_GSS10: AQ214799	1288 431	L34206 AQ214799	<i>Petroselinum crispum</i> protein kinase p34cdc2 (cdc2) mRNA, complete cds. HS_3010_A2_G12_MR CTF Approved Human Genomic Sperm Library D	<i>Petroselinum crispum</i> <i>Homo sapiens</i>	37,816	17-Feb-96
rxa02120	885	GB_PL2: PUMCDC2A GB_GSS10: AQ214799	1288 485	L34206 AQ07596	<i>Petroselinum crispum</i> protein kinase p34cdc2 (cdc2) mRNA, complete cds. HS_5560_B1_H08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate = 1136 Col = 15 Row = P, genomic survey sequence.	<i>Petroselinum crispum</i> <i>Homo sapiens</i>	34,591	18-Sep-98
rxa02126	444	GB_PL2: PUMCDC2A GB_GSS4: AQ0707596	1288 485	AQ0707596	<i>Petroselinum crispum</i> protein kinase p34cdc2 (cdc2) mRNA, complete cds. HS_5560_B1_H08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate = 771 Col = 21 Row = C, genomic survey sequence.	<i>Petroselinum crispum</i> <i>Homo sapiens</i>	36,541 38,482	17-Feb-96 7-Jul-99
		GB_GSS13: AQ494885	411	AQ494885	<i>Homo sapiens</i> genomic clone Plate = 1136 Col = 15 Row = P, genomic survey sequence.	<i>Homo sapiens</i>	40,897	28-Apr-99
		GB_GSS4: AQ707596	485	AQ707596	<i>Homo sapiens</i> genomic clone Plate = 771 Col = 21 Row = C, genomic survey sequence.	<i>Homo sapiens</i>	43,533	7-Jul-99
rxa02148	1266	GB-HTG2: AC007905	100722	AC007905	<i>Homo sapiens</i> chromosome 16q24.3 clone PAC 754F23, *** SEQUENCING IN PROGRESS ***, 33 unorderd pieces.	<i>Homo sapiens</i>	36,051	24-Jun-99
		GB-HTG2: AC007905	100722	AC007905	<i>Homo sapiens</i> chromosome 16q24.3 clone PAC 754F23, *** SEQUENCING IN PROGRESS ***, 33 unorderd pieces.	<i>Homo sapiens</i>	36,051	24-Jun-99
		GB-HTG2: AC007905	100722	AC007905	<i>Homo sapiens</i> chromosome 16q24.3 clone PAC 754F23, *** SEQUENCING IN PROGRESS ***, 33 unorderd pieces.	<i>Homo sapiens</i>	35,402	24-Jun-99
rxa02214	732	GB_GSS13: AQ459868	402	AQ459868	IN PROGRESS ***, 33 unorderd pieces. HS_5116_A1_H04_SP6E RPCI-11 Human Male BAC Library <i>Homo sapiens</i> genomic clone Plate = 692 Col = 7 Row = O, genomic survey sequence.	<i>Homo sapiens</i>	43,035	23-Apr-99
		GB_EST26: AU005050	790	AU005050	AU005050 <i>Bombyx mori</i> p50(Daizo) <i>Bombyx mori</i> cDNA clone ws30188, mRNA sequence.	<i>Bombyx mori</i>	45,902	19-Jan-99

TABLE 4-continued
ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa02316	1137	GB_PL1: F8K7	98581	AC007727	<i>Arabidopsis thaliana</i> chromosome 1 BAC P8K7 sequence, complete sequence.	<i>Arabidopsis thaliana</i>	37,155 29-Jun-99	
rxa02316			600	AJ723424	<i>Haemophilus contortus</i> Intestinal mRNA <i>Haemophilus contortus</i> cDNA clone hcgs49_T7. <i>Haemophilus contortus</i> mRNA sequence.	<i>Haemophilus contortus</i>	35,953 10-Jun-99	
rxa02348	831	GB_PR4: AC000134 GB_STS: AF021124 GB_PL1: AT224957 GB_RO: AF022770	203300	AC000134 575 AF021124 4081 AJ224957 577 AF022770	<i>Homosapiens</i> Chromosome 11q13 BAC Clone 137c7, complete sequence. <i>Arabidopsis thaliana</i> <i>Homosapiens</i> tri nucleotide repeat ctg-68, sequence tagged site. <i>Mus musculus</i> peripheral benzodiazepine receptor associated protein (Pap7) mRNA, partial cds.	<i>Homosapiens</i> <i>Arabidopsis thaliana</i> <i>Mus musculus</i>	37,030 06-MAY-1999 41,913 3-Apr-98 35,627 19-MAY-1998 39,652 24-Sep-97	
rxa02411	972	GB_BA1: AB020624	890	AQ258908	<i>Mus musculus</i> genomic survey sequence.	<i>Oryza sativa</i>	39,515 23-OCT-1998	
rxa02411			1605	AB020624	<i>Corynebacterium glutamicum</i> mutl gene for D-glutamate racemase, complete cds.	<i>Corynebacterium glutamicum</i>	98,868 24-Jul-99	
rxa02448	1212	GB_EST18: AA733776 GB_EST38: AW033449 GB_BA1: AB016258	385	AA733776 612 AW033449 2260 AB016258	<i>Vv0303.1</i> Stratagene mouse skin (##37313). <i>Mus musculus</i> cDNA clone IMAGE: 1210589 S' mRNA sequence. EST277020 tomato callus, TMU <i>Lycopersicon esculentum</i> cDNA clone cLEC28F5, mRNA sequence.	<i>Mus musculus</i>	43,864 7-Jan-98	
rxa02448			553	AW014148	<i>Arthrobacter</i> sp. gene for maleylacetate reductase and hydroxyquinol 1,2-dioxygenase, partial and complete cds.	<i>Arthrobacter</i> sp.	35,620 15-Sep-99	
rxa02448			543	AA432042	IMAGE: 2709487 3' mRNA sequence. 2W8001.1 <i>Saccharomyces cerevisiae</i> cDNA clone IMAGE: 782521 5' similar to WP_ T12A7_1 CEG6433; mRNA sequence.	<i>Arthrobacter</i> sp.	36,522 22-MAY-1997	
rxa02449	1036	GB_BA1: AB016258 GB_BA1: CGPUTP GB_BA1: AB016258	2260	AB016258	<i>Arthrobacter</i> sp. gene for maleylacetate reductase and hydroxyquinol 1,2-dioxygenase, partial and complete cds.	<i>Arthrobacter</i> sp.	44,560 10-Sep-99	
rxa02497	1050	GB_BA2: CGU131224 GB_BA1: MTCY20G9 GB_BA1: SCE7	3791	Y09163	<i>C. glutamicum</i> putP gene.	<i>Corynebacterium glutamicum</i>	39,899 8-Sep-97	
rxa02526	1329	GB_GSS10: AQ240233	2260 AB016258	U016258	<i>Arthrobacter</i> sp. gene for maleylacetate reductase and hydroxyquinol 1,2-dioxygenase, partial and complete cds.	<i>Arthrobacter</i> sp.	70,410 8-Sep-99	
rxa02530		GB_OV: S48556	422	U31224	<i>Corynebacterium glutamicum</i> (npk) genes, partial cds.	<i>Corynebacterium glutamicum</i>	96,445 2-Aug-96	
rxa02530		GB_PR2: HSM801056 GB_PR3: HSJ73D10	37218 Z77162	16911 A1049819	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 25/162. <i>Srenomyces coelicolor</i> cosmid E7.	<i>Mycobacterium tuberculosis</i> <i>Srenomyces coelicolor</i>	59,429 17-Jun-98 39,510 10-MAY-1999	
rxa02530			483	AQ240233	CIT-HSP-2385F9.TR.1 CIT-HSP <i>Homosapiens</i> genomic clone 2385F9, genomic survey sequence.	<i>Homosapiens</i>	42,475 30-Sep-98	
rxa02530			195	S48556	{ tandem repeat P1 monomer} [<i>Cacalia galera</i> = sulfur-crested cockatoo, Genomic, 195 nt].	<i>Cacalia galera</i>	50,515 08-MAY-1993	
rxa02530		GB_PR2: HSM801056 GB_PR3: HSJ73D10	2555 AL117532	97912 AL049651	<i>Homosapiens</i> mRNA; cDNA DKFZp434E192 (from clone DKFZp434E192). Human DNA sequence from clone 753D10 on chromosome 20 Contains genes for SST1R (somatostatin receptor 4) and THBD (thrombomodulin), ESTs, STSs, GSSs and CpG Islands, complete sequence.	<i>Homosapiens</i>	39,116 15-Sep-99	
rxa02530			661	A1782764	EST263643 tomato susceptible, Cornell <i>Lycopersicon esculentum</i> cDNA clone cLES20B10, mRNA sequence.	<i>Lycopersicon esculentum</i>	34,248 23-Nov-99	
rxa02530		GB_EST33: A1782764 GB_GSS9: AQ121479	521	AQ121479	HS_3084_A2_B02_MF CTT Approved Human Genomic Sperm Library D <i>Homosapiens</i> genomic clone Plate = 3084 Col = 4 Row = C, genomic survey sequence.	<i>Homosapiens</i>	38,689 22-Sep-98	

TABLE 4-continued
ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa02535	1278	GB-HTG3: AC008710	146055	AC008710	<i>Homo sapiens</i> chromosome 5 clone C1978SKB_7E3, *** SEQUENCING	<i>Homo sapiens</i>	35,799	3-Aug-99
		GB-HTG3: AC008710	146055	AC008710	<i>Homo sapiens</i> chromosome 5 clone C1978SKB_7E3, *** IN PROGRESS, *** 39 unordered pieces.	<i>Homo sapiens</i>	35,799	3-Aug-99
		GB-HTG3: AC008710	146055	AC008710	<i>Homo sapiens</i> chromosome 5 clone C1978SKB_7E3, *** IN PROGRESS, *** 39 unordered pieces.	<i>Homo sapiens</i>	34,886	3-Aug-99
rxa02603	1119	GB_BA1: MTIV026	23740	AL022076	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 157/162.	<i>Mycobacterium tuberculosis</i>	37,975	24-Jun-99
		GB_IN2: AC005714	177740	AC005714	<i>Drosophila melanogaster</i> , chromosome 2R, region 58D4-58E2, BAC clone BACR48M13, complete sequence.	<i>Drosophila melanogaster</i>	41,226	01-MAY-1999
		GB_EST19: AA775050	218	AA775050	IMAGE: 868514, 3 similar to gb: Y00371_7na1 HEAT SHOCK COGNATE 71 KD PROTEIN (HUMAN); mRNA sequence.	<i>Homo sapiens</i>	40,826	5-Feb-98
rxa02641	1053	GB_BA1: MTICY48	35377	Z74020	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 69/162.	<i>Mycobacterium tuberculosis</i>	62,678	17-Jun-98
rxa02651		GB_BA1: SC4A10	43147	AL109663	<i>Streptomyces coelicolor</i> cosmid 4A10.	<i>Streptomyces coelicolor</i>	39,109	5-Aug-99
		GB_BA1: MLCL458	43839	AL049478	<i>Mycobacterium leprae</i> cosmid L458.	A.3(2)	62,753	27-Aug-99
rxa02674	1575	GB_BA2: PPU96338	5276	U96338	<i>Pseudomonas putida</i> NCIMB 9869 plasmid pRA4000 p-cresol degradative pathway genes, p-hydroxybenzaldehyde dehydrogenase (pchA), p-cresol methylhydroxylase, cytochrome subunit precursor (pchC), unknown (pchX), and p-cresol methylhydroxylase, flavoprotein subunit (pchF) genes, complete cds.	<i>Pseudomonas putida</i>	58,095	13-MAY-1999
		GB_BA1: SCE9	37730	AL049841	<i>Streptomyces coelicolor</i> cosmid E9.	<i>Streptomyces coelicolor</i>	38,544	19-MAY-1999
		GB_BA2: PPU96339	4464	U96339	<i>Pseudomonas putida</i> NCIMB 9869 plasmid pRA500 p-cresol degradative pathway genes, p-hydroxybenzaldehyde dehydrogenase (pchA) gene, partial cds, and p-cresol methylhydroxylase, cytochrome subunit (pchC), unknown (pchX), p-cresol methylhydroxylase, flavoprotein subunit (pchF), protocatechuate-3,4-dioxygenase, beta subunit (pcatF) and protocatechuate-3,4-dioxygenase, alpha subunit (pcatG) genes, complete cds.	<i>Pseudomonas putida</i>	70,588	13-MAY-1999
rxa02702	1581	GB_BA1: AB015023	2291	AB015023	<i>Corynebacterium glutamicum</i> genes for MurC and FisQ, complete cds.	<i>Corynebacterium glutamicum</i>	99,365	6-Feb-99
		GB_BA1: AB003132	4116	AB003132	<i>Corynebacterium glutamicum</i> gene for MurC, FisQ, FisZ, complete cds.	<i>Corynebacterium glutamicum</i>	99,317	4-Aug-97
		GB_BA1: BLFTSZ	5546	Y08964	<i>B. lacifermentum</i> murC, ftsQ or divD & ftsZ genes.	<i>Corynebacterium glutamicum</i>	99,296	08-OCT-1998
rxa02703	1212	GB_BA1: AB015023	2291	AB015023	<i>Corynebacterium glutamicum</i> genes for MurC and FisQ, complete cds.	<i>Corynebacterium glutamicum</i>	97,468	6-Feb-99
		GB_PL2: VFAMACTRA	1879	Y09591	<i>V. faba</i> mRNA for amino acid transporter.	<i>Vicia faba</i>	38,915	02-DEC-1999
		GB_PAT: E05047	966	E05047	DNA encoding recombinant monoglyceride lipase.	Bacillus sp.	37,158	29-Sep-97
rxa02704	1812	GB_BA1: MTICY270	37586	Z95388	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 96/162.	<i>Mycobacterium tuberculosis</i>	37,946	10-Feb-99
		GB_BA2: AE000961	18765	AE000961	<i>Archaeoglobus fulgidus</i> section 146 of 172 of the complete genome.	<i>Archaeoglobus fulgidus</i>	38,521	15-DEC-1997
		GB_BA1: MTICY270	37586	Z95388	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 96/162.	<i>Mycobacterium tuberculosis</i>	37,850	10-Feb-99
rxa02705	1539	GB_PAT: E26124	6911	I26124	Sequence 4 from patent US 5556776.	Unknown.	97,619	07-OCT-1996
		EM_PAT: E11760	6911	E11760	Base sequence of sucrose gene.	<i>Corynebacterium glutamicum</i>	97,619	08-OCT-1997
		GB_BA1: SC4A10	43147	AL109663	Streptomyces coelicolor	(Rel. 52, Created)	37,856	5-Aug-99

TABLE 4-continued

ALIGNMENT RESULTS

ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxa02706	1221	GB_PAT: I26124 EM_PAT: E11760	6911	I26124 6911 E11760	Sequence 4 from patent US 5556776. Base sequence of sucrose gene.	Unknown. <i>Corynebacterium glutamicum</i>	98,605 07-OCT-1996 98,605 08-OCT-1997 (Rel. 52, Created)	
rxa02707	1653	GB_BA1: MTCY770 EM_PAT: E11760	37586	Z95388 6911 E11760	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 96/162. Base sequence of sucrose gene.	<i>Mycobacterium tuberculosis</i> <i>Corynebacterium glutamicum</i>	34,868 10-Feb-99 98,547 08-OCT-1997 (Rel. 52, Created)	
rxa02710	1686	GB_PAT: I26124 GB_BA1: MLCB268 EM_PAT: E11760	6911	I26124 38859 AL022602 6911 E11760	Sequence 4 from patent US 5556776. <i>Mycobacterium leprae</i> cosmid B268. Base sequence of sucrose gene.	Unknown. <i>Mycobacterium leprae</i> <i>Corynebacterium glutamicum</i>	98,547 07-OCT-1996 37,815 27-Aug-99 52,124 08-Oct-1997 (Rel. 52, Created)	
rxa02711	2235	GB_BA2: XCU45994	6911	I26124 515 AQ484169	Sequence 4 from patent US 5556776. RPCI-11-264A12.TV RPCI-11. <i>Homo sapiens</i> genomic clone RPCI-11- 264A12, genomic survey sequence.	<i>Homo sapiens</i>	52,124 07-OCT-1996 40,856 24-Apr-99	
rxa02716	684	GB_BA2: XCU77781	1203	U45994	<i>Xanthomonas campestris</i> pv. <i>campestris</i> insertion sequence IS1404.	<i>Xanthomonas campestris</i> pv. <i>campestris</i>	39,061 29-Jan-99	
rxa02713	1134	GB_BA1: MTCY770 GB_PRI: D31907	4160	U77781	<i>Xanthomonas campestris</i> pv. amaranthicola Xam1 DNA methyltransferase (Xam1M) gene, complete cds; insertion sequence IS1389 and unknown genes.	<i>Xanthomonas campestris</i> pv. amaranthicola	39,551 9-Feb-99	
rxa02715	111353	GB_BA2: AF108355	1222	AF108355	<i>Xanthomonas campestris</i> pv. amaranthicola insertion sequence IS1389-B unknown genes.	<i>Xanthomonas campestris</i> pv. amaranthicola	40,281 09-MAR-1999	
rxa02722	1449	GB_BA1: BLFTSZ GB_BA1: AB003132	37586	Z95388 599 D31907	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 96/162. <i>Homo sapiens</i> gene for zinc regulatory factor, partial cds.	<i>Mycobacterium tuberculosis</i> <i>Homo sapiens</i>	38,669 10-Feb-99 36,326 7-Feb-99	
rxa02723	789	GB_PRI: HSMTFMR GB_PR3: AC002347	3302	X78710	<i>Homo sapiens</i> MTF-1 mRNA for metal-regulatory transcription factor.	<i>Homo sapiens</i>	37,243 1-Aug-94	
		GB_PR3: HS31016	134977	AC002347	<i>Homo sapiens</i> chromosome 17, clone 297N7, complete sequence.	<i>Homo sapiens</i>	36,232 3-Feb-98	
			87942	AL035593	Human DNA sequence from clone 31016 on chromosome 6q22.1-22.3.	<i>Homo sapiens</i>	37,291 23-Nov-99	
		GB_HTG3: AC011509	111353	AC011509	Contains part of a novel gene, ESTs, STSs and GSSs, complete sequence. <i>Homo sapiens</i> chromosome 19 clone CITB-H1_2189E23, ***	<i>Homo sapiens</i>	37,407 07-OCT-1999	
					SEQUENCING IN PROGRESS **, 35 unordered pieces.			
					<i>B. laevis</i> murC, ftsQ or divD & ftsZ genes.	<i>Corynebacterium glutamicum</i>	99,652 08-OCT-1998	
					<i>Corynebacterium glutamicum</i> gene for MurC, FtsQ, FtsZ, complete cds.	<i>Corynebacterium glutamicum</i>	98,535 4-Aug-97	
					<i>Brevibacterium flavum</i> ftsQ gene complete cds.	<i>Corynebacterium glutamicum</i>	97,235 28-Jul-99	
					<i>Corynebacterium glutamicum</i> genes for MurC and FtsQ, complete cds.	<i>Corynebacterium glutamicum</i>	99,113 6-Feb-99	
					<i>B. laevis</i> murC, ftsQ or divD & ftsZ genes.	<i>Corynebacterium glutamicum</i>	99,113 08-OCT-1998	
					<i>Corynebacterium glutamicum</i> gene for MurC, FtsQ, FtsZ, complete cds.	<i>Corynebacterium glutamicum</i>	99,113 4-Aug-97	

TABLE 4-continued

ALIGNMENT RESULTS						
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit
rxa02813	1108	GB-HTG3; AC009658	171795	AC009658	<i>Homo sapiens</i> chromosome 15 clone 344_A_16 map 15, *** SEQUENCING <i>Homo sapiens</i> IN PROGRESS ***, 29 unordered pieces.	34,622 01-OCT-1999
		GB-HTG3; AC009658	171795	AC009658	<i>Homo sapiens</i> chromosome 15 clone 344_A_16 map 15, *** SEQUENCING <i>Homo sapiens</i> IN PROGRESS ***, 29 unordered pieces.	34,622 01-OCT-1999
rxa02820	1411	GB_RO: MMU65079 GB_BA1: BFU64514	2300 3837	U65079 U64514	<i>Mus musculus</i> actin-binding protein (ENC-1) mRNA, complete cds. <i>Bacillus firmus</i> dppABC operon, dipeptide transporter protein dppA gene, partial cds, and dipeptide transporter proteins dppB and dppC genes, complete cds.	<i>Mus musculus</i> <i>Bacillus firmus</i> 35,013 29-Jul-97 36,859 1-Feb-97
		GB_INI: CET04C10 GB_EST35: AI823090	20958 720	Z69885 AI823090	<i>Caenorhabditis elegans</i> cosmid T04C10, complete sequence. L30-94cT3 Ice plant Lambda Uni-Zap XR expression library, 30 hours NaCl treatment <i>Mesembryanthemum crystallinum</i> cDNA clone L30-944 5' similar to 60S ribosomal protein L36 (AC004684)/ <i>Arabidopsis thaliana</i> , mRNA sequence.	<i>Caenorhabditis elegans</i> <i>Mesembryanthemum crystallinum</i> 35,934 2-Sep-99 35,770 21-Jul-99
rxa02828	572	GB_BA1: MTGY10H4 GB_BA1: MTOKIREP	39160 8400	Z80233 X92504	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 2/162. <i>M. tuberculosis</i> origin of replication and genes rmpA, rpmH, dnaA, dnaN, recF.	<i>Mycobacterium tuberculosis</i> <i>Mycobacterium tuberculosis</i> 39,823 17-Jun-98 39,823 26-Aug-97
rxa02839	470	GB_RO: RATENDOGLY GB_BA2: ECOUW89 GB_BA2: AE000477 GB_BA1: ECOPLSB	3906 176195 11314 3865	L37380 U90006 AE000477 K00127	Rat apical endosomal glycoprotein mRNA, complete cds. <i>E. coli</i> chromosomal region from 89.2 to 92.8 minutes. <i>Escherichia coli</i> K-12 MG1655 section 367 of 400 of the complete genome. <i>E. coli</i> plkB and dgk genes coding for sn-glycerol-3-phosphate acyltransferase and diglyceride kinase.	<i>Rattus norvegicus</i> <i>Escherichia coli</i> <i>Escherichia coli</i> <i>Escherichia coli</i> 38,704 20-Apr-95 99,362 17-DEC-1993 99,787 12-Nov-98 33,761 28-Feb-94
rks03218						

[0216]

SEQUENCE LISTING

The patent application contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/sequence.html?DocID=20050191732>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

What is claimed:

1. An isolated nucleic acid molecule from *Corynebacterium glutamicum* encoding an HA protein, or a portion thereof, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
2. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid molecule encodes an HA protein involved in the production of a fine chemical.
3. An isolated *Corynebacterium glutamicum* nucleic acid molecule selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
4. An isolated nucleic acid molecule which encodes a polypeptide sequence selected from the group consisting of those sequences set forth in Appendix B, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
5. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide selected from the group of amino acid sequences consisting of those sequences set forth in Appendix B, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
6. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
7. An isolated nucleic acid molecule comprising a fragment of at least 15 nucleotides of a nucleic acid comprising a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A, provided that the nucleic acid molecule does not consist of any of the F-designated genes set forth in Table 1.
8. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1-7 under stringent conditions.
9. An isolated nucleic acid molecule comprising the nucleic acid molecule of claim 1 or a portion thereof and a nucleotide sequence encoding a heterologous polypeptide.
10. A vector comprising the nucleic acid molecule of claim 1.
11. The vector of claim 10, which is an expression vector.
12. A host cell transfected with the expression vector of claim 11.
13. The host cell of claim 12, wherein said cell is a microorganism.
14. The host cell of claim 13, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.
15. The host cell of claim 12, wherein the expression of said nucleic acid molecule results in the modulation in production of a fine chemical from said cell.
16. The host cell of claim 15, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, polyketides, and enzymes.
17. A method of producing a polypeptide comprising culturing the host cell of claim 12 in an appropriate culture medium to, thereby, produce the polypeptide.
18. An isolated HA polypeptide from *Corynebacterium glutamicum*, or a portion thereof.
19. The polypeptide of claim 18, wherein said polypeptide is involved in the production of a fine chemical production.
20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B, provided that the amino acid sequence is not encoded by any of the F-designated genes set forth in Table 1.
21. An isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B, or a portion thereof, provided that the amino acid sequence is not encoded by any of the F-designated genes set forth in Table 1.
22. The isolated polypeptide of claim 18, further comprising heterologous amino acid sequences.
23. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid selected from the group consisting of those sequences set forth in Appendix A, provided that the nucleic acid molecule does not consist of any of the F-designated nucleic acid molecules set forth in Table 1.
24. An isolated polypeptide comprising an amino acid sequence which is at least 50% homologous to an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B, provided that the amino acid sequence is not encoded by any of the F-designated genes set forth in Table 1.
25. A method for producing a fine chemical, comprising culturing a cell containing a vector of claim 12 such that the fine chemical is produced.

26. The method of claim 25, wherein said method further comprises the step of recovering the fine chemical from said culture.

27. The method of claim 25, wherein said method further comprises the step of transfecting said cell with the vector of claim 11 to result in a cell containing said vector.

28. The method of claim 25, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

29. The method of claim 25, wherein said cell is selected from the group consisting of: *Corynebacterium glutamicum*, *Corynebacterium herculis*, *Corynebacterium lilium*, *Corynebacterium acetoacidophilum*, *Corynebacterium acetoglutamicum*, *Corynebacterium acetophilum*, *Corynebacterium ammoniagenes*, *Corynebacterium fujiiokense*, *Corynebacterium nitrilophilus*, *Brevibacterium ammoniagenes*, *Brevibacterium butanicum*, *Brevibacterium divaricatum*, *Brevibacterium flavum*, *Brevibacterium healii*, *Brevibacterium ketoglutamicum*, *Brevibacterium ketosoreductum*, *Brevibacterium lactofermentum*, *Brevibacterium linens*, *Brevibacterium paraffinolyticum*, and those strains set forth in Table 3.

30. The method of claim 25, wherein expression of the nucleic acid molecule from said vector results in modulation of production of said fine chemical.

31. The method of claim 25, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, polyketides, and enzymes.

32. The method of claim 25, wherein said fine chemical is an amino acid.

33. The method of claim 32, wherein said amino acid is drawn from the group consisting of: lysine, glutamate,

glutamine, alanine, aspartate, glycine, serine, threonine, methionine, cysteine, valine, leucine, isoleucine, arginine, proline, histidine, tyrosine, phenylalanine, and tryptophan.

34. A method for producing a fine chemical, comprising culturing a cell whose genomic DNA has been altered by the inclusion of a nucleic acid molecule of any one of claims 1-9.

35. A method for diagnosing the presence or activity of *Corynebacterium diphtheriae* in a subject, comprising detecting the presence of one or more of the sequences set forth in Appendix A or Appendix B in the subject, provided that the sequences are not or are not encoded by any of the F-designated sequences set forth in Table 1, thereby diagnosing the presence or activity of *Corynebacterium diphtheriae* in the subject.

36. A host cell comprising a nucleic acid molecule selected from the group consisting of the nucleic acid molecules set forth in Appendix A, wherein the nucleic acid molecule is disrupted.

37. A host cell comprising a nucleic acid molecule selected from the group consisting of the nucleic acid molecules set forth in Appendix A, wherein the nucleic acid molecule comprises one or more nucleic acid modifications from the sequence set forth in Appendix A.

38. A host cell comprising a nucleic acid molecule selected from the group consisting of the nucleic acid molecules set forth in Appendix A, wherein the regulatory region of the nucleic acid molecule is modified relative to the wild-type regulatory region of the molecule.

* * * * *